THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XVI | November, 1925 | Number 8

THE NATURE OF INTELLIGENCE

JOHN P. HERRING

Teachers College, Columbia University

The vital things to attempt in studying the nature of intelligence are three:

1. To find out what it is as experience
2. To find out what it is as bodily process, especially in the nervous system
3. To find out what it is as behavior

The first will be left largely to the introspectionist and the second to the mechanist, leaving most of the space for identifying intelligence as behavior.

Intelligence as Experience

The direct study of experience is made by introspection. Since I have never tried to become skilled in this cloistered art, and since I regard it as Thorndike does, I shall leave it wholly to introspectionists. The methods, results and outlook of introspection as used upon the more complex mental processes seem so unimpressive that I have given them only that little study necessary to estimate their relative merits for this problem.

The troubles are these: since no event can be observed a second time by the same or another observer, verification of original data never takes place. If it did, then the accuracy of observation would vary inversely as the square root of the number of observations. Since, therefore, the number of observations of the same experience is never greater than one, this error is always maximal, as well as unmeasured. If we wish to study typical human experiences apart from their individual differences, these differences ought first to be measured and then eliminated. An assumption that individuals

are alike, and that their experiences are, therefore, comparable is especially contrary to probability in the more complex activities which differentiate human beings in their intelligence, and would presuppose not only verification of individual data but also measurement of individual differences. Quantitative method rests upon the possibility of verification and comparison of data. If there are valid reliable results of introspection bearing upon the nature of intelligence, they are not presented in this paper.

Intelligence as Bodily Process

The study of bodily process, especially of neural activity, is more hopeful. Thorndike's hypothesis is that intelligence is a function of the number and complexity of neural bonds between situations and responses, there being, say:

15,000 ± 8,000 in the high grade idiot
250,000 in the average man
850,000 to 1,000,000 in the highest genius.

More than 20,000 bonds may be required in the education necessary to understand the sentence: "If John and Mary come to your home before dinner, bring them to the pond," while more than 200 may be active in hearing and understanding the words. Invention and reasoning are explained as due to the number and complexity of neural bonds, operating in overt and in vestigial trial and error. Studies in the psychology of arithmetic, algebra and other branches confirm the hypothesis. This statement of Thorndike's position is taken largely from a recent manuscript which was read at the 1924 Washington meeting of the American Association for the Advancement of Science and which appeared in The Educational Record for January, 1925.

The circle reflex, also, doubtless plays a pervasive rôle.

Intelligence as Behavior

The rest of this paper will be devoted to intelligence as behavior—partly because the typical philosophy of the measurement movement has been behaviorism, and partly because the control of behavior seems more likely to result from the study of behavior itself, than from the study of neurones, though this is perhaps not certain.

Phases of the Study of Intelligence as Behavior

The study of the nature of intelligence as behavior may have seven phases:

First, the collection of the best opinion about the nature of intelligence, which is analyzed, and if possible integrated.

Second, the collection of instances of intelligent behavior, which are amassed until they include all the types and enough samples of each type. These are standardized in use and scoring. They are intelligence examinations.

Third, the testing of the examinations, which are studied as to their validity (what they measure), their reliability (how consistently they give their results), their objectivity (how much the results depend on change from examiner to examiner) and other characteristics. In the light of facts they are revised, and homogeneous sub-tests are given differential weights.

Fourth, the analysis of the final validated content of intelligence examinations for the light it will throw upon the nature of intelligence as behavior.

Fifth, the study of the relations of intelligence with other factors.

Sixth, the study of intelligence as an influence upon behavior, as means toward control.

Seventh, the problems of intelligence.

Opinions Concerning the Meaning of Intelligence

I have collected the opinions of about 25 men concerning the nature of intelligence. The British and American Symposiums and books by Spearman, Thorndike, Woodworth and Thurstone are the main sources. A few of the more interesting definitions follow with comment.

1. Binet—"Binet's conception of intelligence emphasizes three characteristics of the thought process: (1) Its tendency to take and maintain a definite direction; (2) the capacity to make adaptations for the purpose of attaining a desired end; and (3) the power of auto-criticism." (Terman: "Measurement of Intelligence," p. 45.)

A group of writers not immediately associated with measurement, led by Dewey and Kilpatrick, have given a singular and independent confirmation of Binet's view. They are concerned with the application of philosophy to method in education. Their question, therefore, is: What is intelligent behavior in education? They identify educa-

tion with life. Their answer is that educative behavior arises from needs of the organism and from environmental obstacles; that it proceeds by formulating purposes and plans, by executing plans, and by criticising the whole process. These aspects are said to represent a more intelligent behavior in education than Herbart's five formal steps.

2. Stern—"General adaptability to the new problems and conditions of life."

3. Terman—Intelligence is ability for abstract thinking.

Definitions like Stern's and Terman's, and others to follow also receive confirmation from the Dewey-Kilpatrick view of intelligence as purposing, planning, doing, and criticising. Binet's definition, which Terman approved, seems logically consistent with this one of Terman's.

4. Thorndike—"The power of good responses from the point of view of truth or fact," taken in gross total situations or abstractly; directly experienced or thought of.

Spearman's criticism that we have no criterion for goodness of response would fail if we should later show that there are technical means for the validation of responses.

5. Meumann—General capacity "for independence, originality and productiveness in thinking."

Henman has called this view narrow.

6. Sherrington—"Integrative action of the nervous system."

This may not have been proposed by Sherrington as a definition of intelligence but it is suggestive.

7. Haggerty—Include "sensation, perception, association, memory, imagination, discrimination, judgment and reasoning." Exclude "emotion, instincts, will activities and so-called character traits."

8. Buckingham—Not Ebbinghaus' association, not Wundt's attention, but "the nature of acts which the central capacity, whatever it is, makes possible." And again, "ability to learn."

9. Peterson—Intelligence is "a biological mechanism by which the effects of a complexity of stimuli are brought together and given a somewhat unified effect in behavior. It is a mechanism for adjustment and control and is operated by internal as well as by external stimuli."

10. Pressey—Rather: what will the tests do?

11. Thurstone—In three aspects: "(*a*) the capacity to inhibit an instinctive adjustment, (*b*) the capacity to redefine the inhibited

instinctive adjustment in the light of imaginally experienced trial and error, (*c*) the volitional capacity to realize the modified instinctive adjustment into overt behavior to the advantage of the individual as a social animal."

And again

"Intelligence considered as a mental trait is the capacity to make impulses focal at their early, unfinished stage of formation."

12. Freeman—Structurally: sensory capacity, perceptual recognition, quickness, range, or flexibility of association, facility in imagination, span of steadiness of attention, and quickness or alertness of response.

I have teased apart the elements of all these and of other definitions, and have set the name of each author against all those elements which he makes emphatic. There was no great difficulty in integrating almost all elements, accentuated by these psychologists of the four countries, into the following coherent, composite statement of the nature of intelligence.

Composite Statement of the Meaning of Intelligence

Intelligence is conscious, biological response to stimulus (Thorndike, Watson, Franzen, Peterson), resulting in environmental readjustment (Binet, Stern, Terman, Colvin, Pintner, Thurstone), by conscious solution of problems not before solved by the same animal (Meumann, Stern, Ballard, Pintner). In life process there are frequent conflicts among instincts, ideas and environments (Franzen, Thurstone), individual and social, by which satisfying responses are inhibited (Thurstone) and annoying responses are incited. Trial and error solutions (Thorndike, Thurstone) are sometimes complete acts, sometimes almost infinitesimal behavior commencements, especially in speech mechanisms (Watson). There is always a more or less endurable balance between satisfyers and annoyers in biological mental life. An organism which manages its learning so as maximally to improve the balance is said to approach its limit of educability, that is, its intelligence (Franzen, Dearborn). Intelligence involves millions of neuro-conscious activities which convention has for the time more or less fixed in the following verbal categories: memory (Haggerty), association (Freeman, Haggerty), imagination (Freeman, Thurstone, Haggerty), attention (Freeman), discrimination (Haggerty), judgment (Haggerty), reasoning (Haggerty), analysis (Ballard), integration (Ballard, Watson, Franzen, Peterson).

I have noted in connection with each element in this statement the names of authors, the collection of which is, of course, far from complete. If details of individual opinion are at times presumably blurred, especially through omission of context, but also in other ways, and perhaps with resulting unfairness, there may be compensation in the fact of relative completeness of statement.

The collections of definitions of intelligence might be more exhaustively completed to include more men and a finer search through the writings of each man. It seems likely, however, that those here assembled are already somewhat representative of what man has come to think intelligence means. This view is confirmed by reflecting that the composite statement is not far from what we should expect from the attempt to integrate the main contributions to the meaning of intelligence from normal psychology, from abnormal psychology, from structuralism, from mechanism, and from behaviorism.

But this account of agreement does not sound much like Spearman's report of the general situation. His doctrine is hopeless dissonance and paralysis of measurement pending the establishment of harmony. He presents conflicting views on the organization of intellect as unifocal (Hart and Spearman), as bifocal (Thomson), as multifocal (Thorndike) and as non-focal (Thorndike, earlier). This situation, he says, paralyzes the power of measuring intelligence. But, in answer to Spearman, it may with propriety be debated whether the solution of this problem must precede measurement—whether we must know the organization of intellect before we can measure intelligent behavior—and even whether we must know what intelligence is in any absolute sense, or as experience, or as neural function, before we can measure it either as neural function or as behavior. I think the answer is negative, and I regard the point proved if it is later shown that, however indirectly, we do measure intelligence.

It should not become customary to quote a number of definitions and then forthwith to impute an obvious disparity (Spearman), or a palpable agreement (Terman), leaving it to gentle criticism to propose criteria for judging between likeness and difference. Rereading Spearman, I find some of the jarring elements fitting well enough into the composite. For instance, Binet's wholly reprehensible fluctuation (Spearman's seeming attitude) from intelligence as *memory plus language intelligence plus sensorial intelligence* to his concept four years later of intelligence as *comprehension plus invention plus direction plus censorship* has merely added to the collection of elements to be fitted into the integral view of intelligence.

We cannot, with Spearman and Koffka, properly be held back from measurement because we do not know what we are measuring. Let anyone, who can, propose a means of finding out what intelligence is without resort to quantitative method. If we really have to know what we are measuring at the outset or never, then we shall never know. A main object of the study is to define intelligence verifiably. We cannot tell what gravitation *is*, or what electricity *is*, but we have laws by Newton and Einstein; and we have the volt, the ampere, and the ohm. We should like to be able to identify the type and characters of some *ding-an-sich*, intelligence. We should like to achieve valid comparability of experience from man to man, and so make verified descriptions of consciousness corresponding to intelligence as identified once in the neural machine and again in the stimulus response bonds of behavior. We should like to be able to describe the mechanisms of intelligent acts with great specificity; as for instance, neural impulse passed from right fore-arm pain receptor number 83,517 over neuron number 4,867,487 reaching a synapsis which had been subjected to n molecules of CO_2 and m molecules of alcohol—but this is doubtless reserved.

At least until we can come more nearly to this ideal it does not help as much toward cognizing the nature of intelligence to say, faculty wise, that intelligence is a power of abstract thought, a tendency to make impulses focal, an ability to evoke correlates—except as we use these definitions as points of departure in the amassing of instances of intelligent behavior.

Any search after the nature of intelligence will show itself to have a prospect of success when it aims to trace the whole of the basic psychological circle: stimuli, receptors, afferent neurones, association neurones, efferent neurones, end organs, muscular and glandular motions, stimuli, receptors and so around the circle again. Dethrone the stimulus with Thurstone? Perhaps we should dethrone the neurone, the muscle or the gland? Or with Spearman search out principles of cognition by speculative means, and then, apparently with both Spearman and Koffka, having propounded a definition, keep it unchanged through the thick and thin of measurement?

Instances of Intelligent Behavior

If any one had possessed unquestioned knowledge of the nature of intelligence as behavior, that person should have communicated it to every one making intelligence examinations. Since no such person

appeared, the test creators did approximately what was done in the construction of composition scales. In view of the certainty that not all the competent would emphasize the same aspects of composition and of the likelihood that all of the aspects emphasized would be worth some weight, composition was not defined, except by each judge for himself alone. The result was an averaging of judgments regarding the nature of composition as written behavior, plus an averaging of judgments regarding the excellence of each sample of composition. This was the validation of the composition scale—the guarantee that it was a scale of composition rather than of something else. Composition itself was directly judged.

A similar thing has occurred in the construction of intelligence examinations. Binet, Terman, Thorndike, Whipple, Haggerty, Dearborn, Colvin and many others have amassed instances of intelligent behavior according to their own judgment. The tests they made have this guarantee, severally and at the source, that, however indirectly, they do measure intelligence. Thorndike recently expressed to me orally his confidence in this as a partial method of test validation. To trust it alone is only less naive than wholly to disregard it. As a matter of probability, neither zero weight nor all the weight is especially likely, while as a matter of fact, substantial positive correlations obtain.

Therefore a real advance is made in this amassing of instances by able judges. Abstract definitions are replaced by detailed accounts. Meaning is manifoldly specified, for while definition is mere symbol, instances are things symbolized. Since mere definition does not specify, it cannot be fully understood.

If so much can truly be said for these samplings from the universe of intelligent human behavior, it is a good fortune that many of them have already been made. Perhaps half a million are of record in standardized form—in tests including analogies, completions, opposites, directions, vocabularies, associative shiftings, classifications, and literally scores of others. Perhaps each of 500,000 samples has had the judgment of an author and some of them the systematic judgments of several. While final confidence may not be placed in the quantities of estimate, it is possible that there exist half as many specific examples as there are stimulus-response bonds in the ablest human being. Both the absolute and the relative number of test elements have been multiplied; new types have been added; and invention has proceeded so far as to bring a problem of criteria for distinguish-

ing between them. The number of psychologists who believe they can tell what a test measures either by inspection or by introspection diminishes directly with the spread of technical means to this end. Indeed, we do not completely know what specific traits lie within total intelligence; what overlapping exists among traits within and without total intelligence; what traits, if any, are wholly without intelligence; or in what degree our present sampling is yielding selective refinement and in what degree a narrowing of desirable content through least resistance and repetition. Not knowing whether any particular selection from among existing test materials measures total intelligence in any specific sense, we may perhaps surmise more safely that it does not—at least until we have means for showing that it does. Since, in the meantime, judgments by psychologists have and can have no means of verification, effort is more profitably spent upon the genesis of criteria than upon the composition of opinion. A guarantee that the test makers do actually evaluate while making test problems, lies in the fact that they are often the very persons whose definitions I have quoted. Spearman's assertion that none has put himself to much trouble to show that test content corresponds to definition is even more seriously impugned by Franzen's "Attempts at Test Validation." Frazen computing all the inter-correlations of 14 American intelligence tests, including only those of his own selection, found substantial correlations among them, the average of which I have estimated to be .74 after correction for attenuation. It would be 1 if the tests were identical. This means a substantial amount of agreement among the author judges and at the same time leaves room for the improvement of the validity of one test by the addition of others. While only 57 pupils were examined, the whole pool of 14 tests and 91 inter-correlations has significance. The National, a Terman, an Otis, the Haggerty and the Illinois General correlate with the whole pool from .90 to .93, and higher than any others, and therefore rest, so far as this evidence goes, at the apex of this pyramid of American judgment about the nature of intelligence.

This pyramid is a pyramid indeed; it rests upon tens of thousands of judgments made by American psychologists in the acts of test construction. Upon these judgments rest immediately the test problems assembled and published by these psychologists. Of all these, Franzen selected 14 examinations from his preliminary knowledge of their worth, and upon the intercorrelations of this restricted group rests the selection of the five tests at the apex.

Frazen has accomplished in a refined manner a fundamental synthesis. If one had to propose a means of comparing two views of the meaning of intelligence, one would suggest stating the meanings in words and then multiplying examples of intelligent activity. Comparison of examples would then reveal the degree of similarity of meanings. A coefficient of correlation between two intelligence tests when corrected for chance errors is such a comparison, and indicates, in the case of this very important work of Raymond Franzen, a substantial agreement among certain American students of intelligence as to the meaning of the term. The correlation states that in their ideas of intelligence the judges agree to an extent of 74 per cent in the sense that 74 per cent of the examination material, while usually differing in content or form or both, is a measure of an identical thing. Spearman's charge of disharmony requires a lower correlation. This is proposed as a quantitative answer to Spearman's unquantitative, speculative assertion.

The definition of the content of intelligence by means of a sampling from the universe of imputed intelligent human behavior is represented by the following test pages:

sentence completion
number series completion
picture completion

vocabulary
opposites
synonyms

reading for comprehension
irrevelant word or object in a group
word analogies
form analogies

repeating digits backward
verbal absurdities
picture absurdities
mixed sentences
mixed picture series

How anyone can inspect the important array of problems represented by such sample pages feeling that, by and large, they measure something not closely related to intelligence by almost any of the definitions presented, is a patent mystery! I, for one, am content to mean those pages when I say intelligence—until we get better pages,

and then I shall mean those better pages! The definition is not emptied but filled.

But, however good this situation may now seem, it is by itself insufficient. The analogy between the elements of tests and the situations of life should not be assumed. It is improper to judge that if James knows enough to get another chair when another guest arrives, he will do correspondingly well with pencil and paper verbal analogies like 4 *chairs are to* 5 *chairs as* 4 *guests to* 5 *guests*, or *walk is to Twentieth Century Limited as tom toms to Stokowski*, or *musician is to jazz as God to profanity*. *Big is to little as palace is to hut* may seem to resemble a relatively non-vestigial, concrete situation like *if there are additions to the party, get the larger machine* or a verbal analogy like *five persons are to seven persons as Ford to Rolls Royce.*

Ballard, also, states that absurdities measure the power to construct coherent wholes—which sounds curiously like another faculty, implying fore-knowledge of correlations between absurdity tests and a faculty of constructing coherent whole houses or coherent whole kettles of fudge. We may prejudge the completion test to correlate highly with a faculty to complete everything else in turn—from love affairs to cross-word puzzles. Confirming this criticism come even some among our cousins, the clinical psychologists, who presently insist that while a particular one minute problem may seem to measure caution, we must await quantitative outcomes before we know. Analogies, also, are said by Ballard to measure knowledge of words plus relation solving. The interesting thing is rather a series of regression weights mathematically determined and indicating how much is contributed by knowledge of some particular group of words and how much by solving some particular group of relations. When we can not say how much, often we can not say there is any. Therefore it would be relatively profitless to spend much time analyzing the different types of problems, and expressing judgment about what they measure—not because this is not a necessary thing to do, but because for the present it has been done enough, and also because its purpose is much less to ascertain fact and much rather to disclose suitable speculative positions from which to take immediate departure over a quantitative route to the land of the tree of knowledge of good and evil examinations. I did sav it has been done enough—and it has, except when new tests are under creation.

Since we have defined first, intending afterward to measure the validity of our defining, we have been cutting without enough margin

at either end, for if we ask: How do we know that this definition defines intelligence, we may with equal propriety ask reversely: How do we know that this word intelligence is what the definition defines? Either may be, for the moment, *the* frame of reference, and both frames are moving. There are, however, two reasons for regarding the definition as a safer point of departure than the word: its fullness and its explicitness. We have pushed these qualities even further, filling the definition in turn with all the detail we can, trying to make it a representative sampling of an inclusive inventory of the universe of intelligent human behavior. We may then trust it the more, if after knowing thus what it has in it, we find reason to like it the better. When we finish, we may be surprised at the nature of the content and determine upon some name other than intelligence to symbolize it—and quite without hurt. Or we may decide that intelligence is still the best term.

Both because the analogy may be weak between stimuli which are tests and those which are situations in life apart from tests, and because human judgment must always seek rest upon as much verification as it can find, in order to escape the waste of major reversals in each generation, the examinations must have a rigorous researching quite independent of the considerations already proposed.

Validation of Certain Instances of Intelligent Behavior

In the meantime the instances are printed into examination booklets to be ingested, American plan, by millions. Unfortunately, some authors rest content here, and some critics suppose the contentment to be general. To the credit, however, of measurement, there is a small but persistent group of persons engaged in the difficult work of a more ultimate validation of intelligence examinations.

An early method, employed by Terman and others, was comparison of test scores with success in school. Probably none doubt that academic success depends in generous measure upon ability to handle such situations as are presented in intelligence tests. It has been justifiable to use school success over a broad field of subject-matter as a better criterion of intelligence than marks in any one subject, for, apart from length of tests, educational age, as a whole, usually correlates more closely with mental age than does arithmetic age, or reading age.

This criterion has usually been composed in part of teachers' judgments and age-grade relations, both of which probably resemble

school success except for chance errors. But since education seems also to depend in a notable degree upon two variables in part irrelevant to intelligence, namely, environment and motivation, it is not ultimately right to employ unequally stimulated schooling as a criterion of intelligence—not even as one of the criteria of intelligence. Essentially the same arguments appertain in the case of social achievements in adult life, such as social position, prestige, the number and quality of publications or of scientific laws created.

Why is unevenly motivated social achievement in school not a proper criterion of intelligence? First, because it is not intelligence—because its correlations with intelligence are still far from unity when chance is eliminated and vary from low to high directly with measures of opportunity and effort; and second, because if we identify intelligence with success, we tend to hold individuals to their past rates of achievement instead of introducing improvement. Wiggam indicates a chasm between intelligence and education by saying that intelligence is the thing that enables a man to get along without education while education is the thing that enables him to get along without the use of his intelligence. We should measure intelligence as directly as we can, never being willing to estimate it from factors which imperfectly represent it by means of theoretical corrections of faulty original data. Schooling as commonly found has been an imperfect measure of intelligence.

It is both curious and deplorable that, in the annals of test construction, great attention has been given to the lesser of the two factors and almost none to the other. Franzen has the distinction to have singled out the equation of interest and effort as an element of major significance in problems of validation. Controversies by Colvin, Bagley, Terman, and, alas! Lippman, have raged about the question of individual inequalities of exposure or opportunity, while none has ever genuinely equated them and the argument steadily refused to concern itself with the more important issue of the control of motivation. The very excellence of the practical work with accomplishment ratios, occasionally done in these days before genuine validation, is none the less a hopeful sign.

With the relative failure of the usual schooling as a criterion in our researches, three recourses are open: first, to obtain groups of examinees in which environment and motivation are highly and therefore presumably equally favorable; second, to select from larger groups individuals who have had optimal environment and maximal

motivation in their work over a period of years; and, third, to eliminate from the results, by means of partial correlations, the effect of inequalities of opportunity and effort. These three methods, in some degree independent of each other, should all be employed. Since they may constitute the crux of the behavioristic definition of intelligence, and since they may be the most direct and sound methods of validating intelligence tests, they are worth detailed consideration.

I propose, therefore, three somewhat distinct plans for testing the instances of intelligent human behavior, the first of which has been proposed before by Franzen.

Plan No. 1. Solution through Motivation

Validation by the first of these methods is somewhat onerous, but not without compensation. It means time and money, but also fundamental immediate benefit to children used in the study. There is required, say, 2500 pupils, ranging somewhat evenly through ages 8 to 13, with no selection in these age groups. These children must be ideally environed, taught, and actuated, so that they will be responding to educative stimuli with maximal contribution. These conditions must continue for a number of years until we have in these age groups only children who have had only this effective education. The undertaking, justifiable therefore even as education merely, should make its appeal especially to administrators who think research good. Measures of equality of opportunity and of equality of interested effort must be made at intervals until a period is reached when both opportunity and effort, ceasing to improve in response to any known stimulus, are in this sense maximal. Since, however, in practice children never would all become just maximally actuated, the solution is hindered—if it is regarded, rightly, as aimed toward the isolation and identification of intelligence and nothing but intelligence. When a child has a difference between his intellectual ability and his educational achievement, attempts to diagnose and prescribe will not always be perfectly successful, even after a period of years. The measures will contain a factor, intelligence, plus a factor, chance, plus a miscellaneous factor, x. x stands for the total of factors, not intelligence, and not chance, which keep the correlation between intelligence and education below one, and may be called opportunity plus motivation. x presumably contains factors which deter children from doing their best, and which are difficult to identify and to treat remedially. If, with administrative, supervisory, instructional and scientific

effort, the magnitude of x approaches a maximum, then the correlation between mental age and educational age approaches unity when corrected for chance errors of measurement; and conversely, if the corrected correlation between mental age and educational age approaches unity, then x approaches a maximum. In other words, when a child's education is proceeding with perfect effectiveness, he is utilizing what intelligence he has and his achievement corresponds to his ability to achieve. This is the fundamental hypothesis underlying this experiment in education and this search into the nature of intelligence.

Parallel with the experimental group must run a control group of equal numbers, similarly selected in all points, and not given an ideal environment with maximal actuation. The control group will represent conventional methods of education which will no doubt give us just what we want: motivation left relatively to chance and exhibiting large inequalities from child to child.

At the outset, and at intervals, intelligence and educational examinations are administered, by means of which the experimenters will review the situation for evidence that the children are maximally actuated. The evidence will be of four sorts: first, the corrected correlation between mental age and educational age will approach unity; second, the differences between Stanford mental age and Stanford educational age will approach zero; and, third, it will be increasingly difficult to tell which pupils are working with most zeal, because the standard deviation of measures of zeal will approach zero; and, fourth, the differences between the experimental and the control group will tend to become constant. The inequalities of environmental opportunity and of effort will have been eliminated; each child will be doing his best work in each field, so far as human teaching and effort could accomplish this.

At this moment, which should occur early in the program, the correlation between educational scores and intelligence scores, when corrected for chance errors, will be either one or less than one. If it is one, then we have directly verified a part of the composite definition of intelligence: namely, that intelligence is the limit of educability. If it is less than one, it is possible that we can make it one by selecting for the different tests those weights which make, by means of least squares, the highest correlation between the intelligence examination as a whole and the educational scores or ages at their limit. If, after we have the best weights for the different tests, it is still less

than one, then it may be that motivation was unequal from child to child, or that tests were not inclusive enough or valid enough, or that the theory that intelligence is, among other things, the limit of educability, is wrong. If we do find a battery of tests which will predict, by means of regression equations, the limit of educability before it is reached, then the whole program of the accomplishment ratio or accomplishment difference follows. Conversely, that we seem able already to use accomplishment ratios effectively for individual control may be prophetic of the outcome of such an experiment.

It is, of course, possible that man can never master techniques for the motivation of children's learning sufficiently for the experiment as outlined. If so, then we still have other recourses within the limits of the data already proposed.

Plan No. 2. Solution through Selection

Like the first one, Plan No. 2 is direct in method, not calling for relatively circuitous means to rectify inadequacies of original data. It is not right to rely upon mathematics and other less direct methods of estimate unless the more immediate approach fails. Because reason is fallible, man must always prefer the more direct logical routes. Every effort is to be made, therefore, for the sake of the children of the experiment, as well as of the progress of science, to master methods of motivation. Assuming, however, that these will remain for some time even more fallible than man's intelligence as applied to scientific inquiry, we may still attempt a solution.

We may select from each age-group those individuals who are most actuated and proceed as before with prediction by means of regression equations. This process is possibly the simplest of the sound techniques for the validation of intelligence tests, and the soundest of the simple techniques. It is applicable to any school group, whether heartily at work or not; of the school system it requires no reorganization—nothing but availability for such testing as is indispensable in running a good school—the administration of intelligence and educational examinations; it yields results immediately and basically useful to the school in the form of accomplishment ratios; it bases its criterion upon those pupils who are maximally actuated, not calling often upon mathematics to make up for inadequacies of original data; it measures the limit of educability for each pupil, whether it defines intelligence or not; and it is as free from the expense and trouble of experimental motivation as from the length and complexity of mathe-

matical computations which lose readers so magically by the wayside and which at times, in pathetic contrast with the fool and his money, separate even the competent from his conclusion. Witness Spearman's failure in interpreting his correlations between columns of correlations, even in a statistical context which he calls simple.

There is however still another recourse.

Plan No. 3. Solution through Partial Correlation

Instead of experimentally eliminating, by Plan 1 or by Plan 2, the differences resulting from inequality of opportunity and effort, these are measured and removed from the numerical results by means of mathematical formulæ involving the regression of educational age upon mental age, apart from the influence of opportunity and effort. A difficulty may appear in getting good measures of x, *i.e.*, opportunity plus effort. These do not need to be measured separately and may be taken as functions of educational age and mental age—perhaps educational age minus mental age. From this procedure, we ought to get the same weights and norms for our intelligence tests as if we had succeeded in actual equation of opportunity and effort in all the children. In this we are not without an empirical test. We shall be satisfied, provided the weights and norms obtained from the data of several groups, which are alike except that they are in different degrees actuated, approach each other as the degree of actuation approaches a maximum. Even if we find feasible both first and second plans, which actually eliminate extraneous factors, we shall still wish to apply this test of consistency. All three plans must be used, now with common examinees and now with different examinees.

There remains that indispensable test of all science, the repetition of the experiment by other students.

As the outcome of the whole affair thus far we should have measures of intelligence, based upon the integrated judgment of many psychologists, statistically proven to have much inner coherence, and in human nature itself tested as the limit of successful intellectual effort when stimuli within and without are the same for all. These measures can be made to yield reliability coefficients of .99 in unselected age groups and probable errors of two mental months.

Intelligence is what certain tests now measure; presently we will improve the criteria and the tests. Later intelligence will be what they then measure. There never will be any Kantian *ding-an-sich*, never any absolute, finally defined thing, concerning the name of

which, with its elusive and nominal grandeur we care relatively little, but always a something which we have imperfectly isolated, and which we can use for specific purposes with a known and small average error. The measurement of total intelligence seems feasible.

I have not prepared a discussion of most of the important aspects of the nature of intelligence; the analysis of intelligence, which would take us into focality, mental types, mechanical, social and abstract intelligence, verbal and non-verbal mindedness; or the relations of intelligence with other factors, which would involve us with emotions, feelings, conations and character; or the influence of intelligence, which would introduce pragmatism, control and program; or the problems of intelligence, which would require for each the designation of appropriate experimental and statistical techniques and an estimate of possible outcomes.

I invite criticism particularly to 7 points:

1. Introspection gives no very lively hope of revealing in a useful sense the nature of intelligence.

2. The best criterion of intelligence suggested herein involves equalized opportunity, and equalized effort and interest.

3. The second best criterion involves regression equations for estimating what the results would be if opportunity and effort were equalized. It would be interesting to use both, so as either to verify the statistical method for doubting Thomases or to invalidate it for credulous Johns.

4. The statistically combined judgments of many psychologists is worth much for test validation.

5. The following are not rigorous criteria: school marks, age-grade relation, teachers' judgments of intelligence, adult social achievements.

6. Validation by means of correlation with a criterion comprising two groups, say of college professors and of Vineland inmates, falls under the ban as a special case of validation by means of social achievements.

7. There should soon be a careful, long-time research combining experimental approach through motivation and selection with statistical analysis.

THE READING TESTS OF THE STANFORD ACHIEVEMENT SCALE AND OTHER VARIABLES

FLORENCE L. GOODENOUGH

Child Guidance Clinic, Minneapolis

The reading test of the Stanford Achievement Scale is made up of three parts "intended to cover the three main factors involved in getting the thought from the printed page; knowledge of word meaning, ability to comprehend relatively short language units (sentences), ability to grasp the meaning of larger units (paragraphs). These factors are closely interrelated but they are not identical. They make different demands on the span of attention and on the ability to grasp the significance of a language unit as a whole, as contrasted with the ability to understand words, phrases, or other isolated fragments."[1]

In the light of the statement just quoted, it was felt that it would be of interest to examine both the reading test as a whole and its separate parts with reference to (*a*) correlation with general intelligence as measured by the Stanford-Binet Scale, (*b*) correlation with the vocabulary test of the Stanford-Binet Scale, and (*c*) the relative importance of chronological age, grade placement, and mental age in reading ability. Incidentally a number of significant facts with specific bearing upon the particular group studied have been brought out in the course of the investigation.

The data upon which this study is based are taken from the records of 100 children who were studied at the Minneapolis Child Guidance Clinic between October, 1924 and April, 1925. This clinic is established for the study and treatment of children who, in one way or another, are failing to make satisfactory adjustments in their social group. Specifically, their difficulties range from such definitely anti-social acts as stealing, persistent truancy, running away from home, sex offenses, and the like, to lack of social adaptability, food faddism, minor disobedience in home or school, temper tantrums, etc. The majority of these cases were referred to the clinic either by the home or by the school; but a few were sent in by the Juvenile Court, others by social agencies such as the Children's Protective Society, and a small number were examined by request of the family physician. A consistent attempt has been made to confine the clinic

[1] Kelly, T. L., Ruch, G. M., and Terman, L. M.: "Stanford Achievement Test, Manual of Directions." World Book Co., Yonkers, 1922. See p. 8.

services to children of normal intelligence, and while it has not been possible to exclude all cases of mental deficiency, the number of such children handled has unquestionably been much smaller than would otherwise have been the case. The distribution of cases, according to school grade, chronological age, and Stanford-Binet IQ is shown in the following tables.

TABLE I.—AGE DISTRIBUTION

6	7	8	9	10	11	12	13	14	15	16	17
1	3	12	13	15	12	8	15	10	8	2	1

Total cases 100. Mean age 11.7. SD, 2.48 years.

TABLE II.—GRADE DISTRIBUTION

2B	2A	3B	3A	4B	4A	5B	5A	6B	6A	7B	7A	8B	8A	9B	9A
2	5	9	8	6	7	7	11	7	6	6	5	12	6	2	1

Total cases 100. Mean grade 5.83. SD, 3.95 half years.

TABLE III.—INTELLIGENCE QUOTIENT DISTRIBUTION

60–69	70–79	80–89	90–99	100–09	110–19	120–29	130–39	140–49
5	11	20	18	23	13	5	1	4

Total cases 100. Mean IQ 97.6. SD, 18.3.

By inspection of Table III it will be seen that in intelligence these children do not differ greatly from a random sampling of the school population. The mean IQ is but slightly below 100, and while the variability of the group is distinctly greater than that reported by Terman[1] for the 905 children used in the original standardation of the Stanford-Binet Scale, it agrees closely with Dickson's[2] findings for first grade children in California, and with Wilson's[3] report for unselected 12-year-olds in New Jersey. Dickson's tables show a standard deviation of 19.7 points of IQ, Wilson's 17.9 points, while for the group under consideration the corresponding figure is 18.3 points.

Intercorrelations have been calculated for the following variables: Chronological age, school grade, mental age, mental age calculated on the basis of the remaining tests in the scale when the vocabulary

[1] Terman, L. M.: "The Measurement of Intelligence." Houghton Mifflin Co., 1916.

[2] Dickson, Virgil: The Relationship of Mental Testing to School Administration, with Special Reference to Children Entering School. Master's Thesis, Stanford University.

[3] Wilson, G. M.: Standard Deviations of Age Scores and Quotients in Typical Groups. *Journal of Educational Psychology*, Vol. 16, 1925, 193–207.

test is omitted, vocabulary score, tests 1, 2, and 3 of the Stanford Achievement Battery, and the sum of these three tests which constitutes the total reading score. The range of talent included is shown in Table IV. For comparison with the mean score obtained by this group, Kelley's[1] norms for unselected school children of chronological age corresponding to the mean mental age of these children are also given. It will be seen that in reading ability they average approximately up to the standard for their mentality, but that their mean grade placement is almost a half year in advance of that which is justified either by their mental age or by their reading ability.

TABLE IV.—TOTAL RANGE, MEANS, AND STANDARD DEVIATIONS FOR ALL VARIABLES CONSIDERED

(Kelley 1924 norms are given for comparison)

	Grade	Chronological age	MA	MA vocabulary	Vocabulary	Test 1	Test 2	Test 3	Total reading
Range	2B–9A	6–17	6–17	6–17	10–73	0–114	0–75	0–81	3–261
SD	3.95	2.48	2.27	2.39	11.65	24.57	19.20	19.40	61.65
Mean	5.83	11.70	11.30	11.50	33.20	54.59	34.60	38.70	127.45
Kelley Norm	5. 4	.....	11.3	.....	.....	56	35	38	129

Table V shows the intercorrelations of the various tests for the entire range of talent included. Tables VI, VII, VIII, and IX show

TABLE V.—INTERCORRELATIONS, ZERO ORDER

	Grade	Chronological age	MA	MA vocabulary	Vocabulary	Test 1	Test 2	Test 3	Total reading
Grade	....	.858	.749	.756	.724	.782	.773	.809	.814
Chronological age	.858	....	.604	.605	.534	.579	.619	.634	.627
MA	.749	.604	....	.986	.861	.850	.787	.790	.839
MA vocabulary	.756	.605	.986	....	.842	.845	.779	.780	.830
Vocabulary	.724	.534	.861	.842	....	.832	.796	.836	.852
Test 1	.782	.579	.850	.845	.832	....	.868	.917	.967
Test 2	.773	.619	.787	.779	.796	.868	....	.890	.944
Test 3	.809	.634	.790	.780	.836	.917	.890	....	.976
Total reading	.814	.627	.839	.830	.852	.967	.944	.976	

[1] Kelley, T. L.: "Revised Age and Grade Norms, Stanford Achievement Test." April, 1924.

TABLE VI.—PARTIAL CORRELATIONS, FIRST ORDER, WITH CHRONOLOGICAL AGE CONSTANT

	Grade	MA	MA vocabulary	Vocabulary	Test 1	Test 2	Test 3	Total reading
Grade	....	.564	.580	.613	.682	.601	.668	.691
MA	.564	....	.978	.800	.770	.660	.661	.741
MA vocabulary	.580	.978	....	.772	.763	.647	.616	.727
Vocabulary	.613	.800	.772	....	.759	.702	.728	.786
Test 1	.682	.770	.763	.759	...	.797	.873	.951
Test 2	.601	.660	.647	.702	.797	....	.783	.909
Test 3	.668	.661	.616	.728	.873	.783	....	.961
Total reading	.691	.741	.727	.786	.951	.909	.961	....

TABLE VII.—PARTIAL CORRELATIONS, FIRST ORDER, WITH GRADE CONSTANT

	Chronological age	MA	MA vocabulary	Vocabulary	Test 1	Test 2	Test 3	Total reading
Chronological age	.....	−.114	−.130	−.246	−.288	−.136	−.199	−.240
MA	−.114	.....	.967	.697	.640	.495	.472	.595
MA vocabulary	−.130	.967	.....	.652	.622	.469	.437	.564
Vocabulary	−.246	.697	.652	.....	.618	.540	.617	.655
Test 1	−.288	.640	.622	.618	.....	.678	.776	.913
Test 2	−.136	.495	.469	.540	.678	.....	.710	.855
Test 3	−.199	.472	.437	.617	.776	.710	.....	.929
Total reading	−.240	.595	.564	.655	.913	.855	.929	

TABLE VIII.—PARTIAL CORRELATIONS, FIRST ORDER, WITH MENTAL AGE CONSTANT

	Grade	Chronological age	MA vocabulary	Vocabulary	Test 1	Test 2	Test 3	Total reading
Grade	....	.768	.158	.234	.417	.449	.535	.515
Chronological age	.768	....	.071	.034	.156	.292	.321	.277
MA vocabulary	.158	.071	.....	−.082	.159	.029	.010	.030
Vocabulary	.234	.034	−.082	.....	.374	.377	.499	.468
Test 1	.417	.156	.159	.374	....	.613	.761	.887
Test 2	.449	.292	.290	.377	.613	....	.709	.845
Test 3	.535	.321	.010	.499	.761	.709	....	.939
Total reading	.515	.277	.030	.468	.887	.845	.939	....

TABLE IX.—PARTIAL CORRELATIONS, FIRST ORDER, WITH TOTAL READING SCORE CONSTANT

	Grade	Chronological age	MA	MA vocabulary	Vocabulary
Grade	....	.768	.209	.248	.100
Chronological age	.768	....	.184	.195	.000
MA	.209	.184	....	.954	.513
MA vocabulary	.248	.195	.954	....	.461
Vocabulary	.100	.000	.513	.461	....

the intercorrelations which obtain when the separate effects of chronological age, school grade, mental age, and total reading score have each been rendered constant by partial correlation treatment.

A number of investigators have called attention to the prevailing tendency among school authorities to promote children on the basis of age rather than ability. It is therefore interesting to note that for the total group (Table V) the correlation between grade placement and chronological age is higher than that between grade placement and either mental age or reading ability. When school grade is rendered constant (Table VII) a negative correlation is found between chronological age and each type of ability considered. In the light of these findings for a group of "problem" children, one is justified in raising the question whether promotion according to age really tends to prevent social maladjustment, as its advocates have claimed, at least when no special provision is made for variation in individual capacity. According to Kelley 1924 norms, only 15 of the 100 children studied are graded correctly on the basis of reading ability, while if we use the educational ages given in the same table as mental age standards, only 15 are graded correctly according to mental age. Of these, 7 are graded correctly by both standards; the remaining cases are graded correctly by one standard but not by the other.

The high correlation of the vocabulary score with the total mental age rating on the Stanford-Binet has been noted by Terman[1] who reports a correlation of .91 between mental age and vocabulary score for a group of California children. The slightly lower figure for

[1] Terman, L. M.: "The Intelligence of School Children." Houghton Mifflin Co., 1919.

this group (.861) can be readily accounted for by the more restricted range of talent (6 yr. 8 mo. to 17 yr. 3 mo. as compared with 5 yr. 6 mo. to 19 yr. 6 mo. for Terman's group). When the mental age is calculated on the basis of the other tests in the scale, omitting the vocabulary test itself, the correlation with vocabulary is but slightly lowered, .842 for the entire range or .772 when chronological age is rendered constant. It would appear, moreover, that for this group at least, the vocabulary score alone affords as satisfactory a basis for the prediction of reading ability as the mental age itself. This is brought out most clearly in Table VIII, where it is shown that when mental age is rendered constant, the vocabulary score still shows a correlation of .468 with the total reading score, whereas the mental age with vocabulary omitted shows practically a zero correlation (0.30) with the reading total.

It was anticipated that Test 3 (word meaning) would correlate more highly with Stanford-Binet vocabulary than either of the other two tests of the reading scale. This appears to be the case, although the fact is not at first apparent because of the spurious element introduced by the very high correlation of both the vocabulary test and Test 1 of the reading scale with mental age, which results in a seemingly equal correlation of both Test 1 and Test 3 with the vocabulary score. When mental age is rendered constant, however (Table VIII), it is found that the correlation of Test 3 with the vocabulary score is .499 as compared with .377 for Test 2, and .374 for Test 1.

It was thought that it would be of interest to ascertain whether the three tests of the reading scale show any difference in order of difficulty for children of different intellectual levels. The one hundred cases studied were accordingly divided into three groups according to intelligence quotient (mental age being disregarded), and the rank order of the educational age ratings earned on the three tests were determined for each individual. The results are shown in Table X.

From the data given in Table X it would appear that bright children excel in dealing with the larger language units, while backward children are comparatively inept in handling material of this kind, but make their best showing in the less complex test of word meaning. This, of course, does not mean that backward children do better than bright children in the word meaning test, but simply that the superiority of the latter is most marked in the case of the paragraph meaning test, while backward children are at somewhat less of a disadvantage in the word meaning test than in those which employ

larger units of thought. It follows, therefore, that if for a given individual the educational age ratings arrange themselves in an inverse order as compared to the complexity of the language units, some suspicion of mental backwardness is justified, while the reverse condition is more likely to be associated with mental superiority.

TABLE X.—COMPARISON OF THE ORDER OF DIFFICULTY OF THE READING TESTS FOR CHILDREN OF DIFFERENT INTELLECTUAL LEVELS

	Intelligence quotient					
	Below 90 Total 36		90–109 Total 41		110 or above Total 23	
	N	Per cent	N	Per cent	N	Per cent
Rating on Test 1 higher than on Test 2	17	47.2	22	53.7	15	65.2
Rating on Test 1 higher than on Test 3	12	33.3	22	53.7	16	69.6
Rating on Test 2 higher than on Test 3	11	30.6	20	48.8	13	56.5
Order of difficulty 1, 2, 3, 1 being easiest	4	11.1	6	14.6	6	26.1
Order of difficulty 3, 2, 1, 3 being easiest	12	33.3	4	9.8	1	4.3

In clinical work especially, it not infrequently happens that the time available for the examination of a child is quite limited. When this is the case and a brief test of reading ability is required, Test 3, for which the time limit is only ten minutes, seems to give a very close approximation to the result which would be obtained by using the entire series of three tests. Reference to the foregoing tables will show that the correlation between the score on this test with the total reading score remains well above .9 even when chronological age, school grade, and mental age are each in turn rendered constant by partial correlation treatment. As a further check on the extent of the error introduced by the use of this single test in place of the total series, the actual differences between the age ratings earned on this test and those obtained from the total scale were tabulated for the 100 children included in this study. It was found that in 67 cases the error amounted to less than one-half year, in 89 cases it was less than a year,

and in only a single instance was it as great as two years. When it is remembered that within a single class, differences of from three to four years in reading ability are not uncommon, it will be seen that the likelihood of grossly misclassifying a child, even by this brief method of testing, is quite inconsiderable. A similar tabulation was made for Test 1. It was found that in 66 per cent of the cases the error amounted to less than a half year, in 91 per cent, to less than a year, while 2 per cent would have been misrated to the extent of two or more years. However the time required for this test is double that for Test 3. Test 1 shows a higher correlation with intelligence than does Test 3, and for this reason it probably affords a more satisfactory basis for prognosis of future reading accomplishment. If, then, a brief reading test is required in cases where no intelligence test has been given, Test 1 will presumably be found more useful than Test 3, but if intelligence test results are available, Test 3 seems to give a somewhat better measure of present reading accomplishment than Test 1. It should be remembered that backward children are likely to be rated slightly too high by this test, and that the ratings of superior children may be somewhat too low. Test 2 shows relatively lower correlations with the reading total throughout, hence the results from this test have not been tabulated separately.

Conclusions

1. A study of the grade placement of 100 school children referred to a child guidance clinic for behavior difficulties, shows that only 15 per cent were graded correctly from the standpoint of mental age, and only 15 per cent were in the grade corresponding most nearly to their reading ability. A constant tendency toward over-grading was observed, which amounted to an average of nearly one-half year, or about the equivalent of one school grade.

2. The correlation between grade placement and chronological age was found to be higher than that between grade placement and any of the measures of ability considered.

3. A high correlation was found between the reading tests of the Stanford Achievement Scale and Stanford-Binet mental age. Of the reading tests, Test 1 (paragraph meaning) correlates more highly with mental age than either the sentence meaning or the word meaning tests taken singly, or the total score on the three tests.

4. The Stanford-Binet vocabulary test shows a very high correlation both with mental age and with the score on the reading tests.

Of the three tests of the reading scale, it correlates most highly with Test 3 (word meaning). The correlation between the total reading score and the vocabulary score is somewhat higher than that between the reading score and the total mental age.

5. Children of superior mental ability tend to make their highest scores on those reading tests which involve the larger language units. They are likely to make their highest educational age rating on the test of paragraph meaning, their lowest on the word meaning test. With backward children the reverse tendency is noted.

6. In cases where economy of time becomes necessary, the use of Test 3 alone as a measure of reading ability introduces a relatively small element of error when compared with the results which would have been obtained by the use of the total series of three tests.

THE EFFECT OF PRACTICE WITH AND WITHOUT KNOWLEDGE OF RESULTS IN GRADING HANDWRITING

A. R. GILLILAND

Northwestern University

The purpose of this article is to report the results of an experimental study of three closely related problems in the grading of handwriting. These problems in turn are the following:

(1) The effect of practice in regarding samples of handwriting by the usual method of grading and without knowledge of the results, that is, without memory of the previous grades given the samples from one grading to the next and without knowledge as to whether there was a tendency to improve, to become worse or remain no better or worse than at the first grading.

(2) The effect of practice in grading handwriting by the use of one of the well known handwriting scales without memory of the grades given any sample from one grading to another, and without knowledge of the amount of error or tendency toward greater or lesser accuracy in the regradings.

(3) The effect of practice in grading handwriting by the use of a scale followed by a period of practice in comparing the grades given the samples with the true grades of the samples as a basis for determining sources of error in grading.

The gradings were made by a small group of junior and senior college students, in a class in educational measurements. Some of these students, as practice teachers, had taught an hour per day for one semester in an elementary school. The others were without teaching experience. None of the class had had previous experience in grading handwriting.

Ten samples of handwriting were selected from the set of practice material given in "Teachers' Estimates of the Quality of Handwriting" by E. L. Thorndike.[1] This monograph contains 50 specimens of handwriting. The true grades of these samples scored on the 0–18 basis of the Thorndike scale and also on the traditional basis of 0–100 are given in the monograph, but these true values of the specimens were not known to the graders until the close of the gradings. Three

[1] See Teachers College Record, Vol. XV, Nov., 1914. See also E. L. Thorndike: Handwriting Teachers College, Columbia University, 1915.

of the samples selected rated high on the Thorndike scale and 4 ranked low. The other 3 ranked between these extremes. Different forms and sizes of handwriting were included in the samples that were selected.

Without having seen copies of any of the standard handwriting scales and without knowing the true scores of these 10 specimens, each member of the class was asked to grade each of the 10 samples. The class was given no further directions than to grade the samples as if they were teachers grading papers in handwriting with 100 as a perfect score. At the next meeting of the class, 2 days later, these same samples were graded a second time without reference to a scale or to the grades previously given. At the next meeting of the class the same samples were graded a third time, again without reference to a scale or to the grades previously given.

The 10 samples were graded a fourth time at the next meeting of the class. The Thorndike scale for Handwriting of Children in Grades V to VIII was used as the basis for this fourth grading. A fifth grading was made with the use of the scale 2 days later. Sometime between the fourth and fifth gradings each member of the class practiced grading 30 other samples taken from the Thorndike specimens. After these 30 samples were graded the true scores of each of these samples were looked up and the samples were then compared with the scale in the light of their known scale values. By this method the students were able to check up on the amount, direction, and nature of their errors in grading. In the fifth grading of the 10 original samples the students had the benefit of this practice in grading with a knowledge of errors in grading. Eight students repeated the practice in grading the 30 specimens and rechecked their errors. They then graded the 10 original samples a sixth time.

The students reported that the time interval of two or three days between each grading period was sufficiently long enough for memory to play no part from one time to another in the grades given. A comparison of the grades given any sample gives ample proof that this was true. There was a possible exception to this in the case of one of the better samples. This sample was distinctive and was probably remembered from one grading to another by some of the graders. Therefore, the effects of the practice were general rather than specific; that is, the improvement in grading would have been practically the same whether these same samples were scored each time or whether different samples had been used each time.

Each score given to the 10 samples by each of the students was compared with the true value of these samples and the deviations computed. The average deviations are shown in the accompanying table. They were made the basis of the study of the improvement in grading with and without knowledge of results. The first three gradings were without the use of the scale or other objective standard measurement. The fourth grading shows the effect of the use of an objective standard without practice in the use of the standard. The fifth grading shows the amount of improvement resulting from about an hour's practice in the use of a scale with a knowledge of the nature and amount of error in grading other samples of handwriting. The sixth grading shows the effects of further practice in grading with a knowledge of results.

The first gradings by the 12 students deviated from the true grades of the 10 samples on an average of 10.9 per cent. The greatest average deviation of any grader was 21.2 and the smallest was 4.2. Although not shown by the table some of the men consistently graded too high, some too low, and others gave some grades too high and some too low. The second grading resulted on the average in a slight improvement in the grading. The deviation was reduced to 8.9. The third grading was not quite so good as the second but better than the first.

	Gradings without scale			With scale	After practice With known score values	
Student	1	2	3	4	5	6
W. G. W.	11.7	5.8	9.3	5.8	4.6	2.1
C. W. S.	9.0	6.6	7.3	3.5	1.2	1.2
A. W. F.	12.2	14.5	8.5	4.8	3.7	3.1
J. P. C.	7.9	11.5	....	3.8	4.2	4.8
E. A. S.	21.2	16.2	16.2	9.0	7.4	....
J. W. S.	10.5	11.9	13.4	7.1	5.3	....
T. I. M.	4.2	4.1	4.2	5.9	1.1	2.0
N. W.	11.2	9.2	10.0	5.9	1.1	1.6
W. W. R.	14.9	7.3	10.2	5.6	3.1	1.2
C. P. G.	7.0	4.0	7.0	5.2	5.9	5.1
J. N. Z.	9.4	6.6	11.1	4.7	4.6	....
E. E. H.	11.7	9.9	10.3	6.3	4.1	....
Average	10.9	8.9	9.8	5.6	4.0	2.6

Showing the average amount of deviation of grades given by 12 students from the true score values in grading 10 samples of handwriting when graded on the 0 to 100 basis. The first three columns are the average errors when the grading was made without the use of a scale. The fourth column gives the average errors with the first use of a handwriting scale. The fifth and sixth columns show the average errors after practice in grading samples with known score values.

These results seem to indicate that there was a very slight improvement in grading without knowledge of the results and without the use of objective standards. This was no doubt due to general improvement in technique in grading and to the development of a habit of giving attention to grading specimens of handwriting. But the improvement is slight and agrees with the results of other studies in practice without knowledge of results.

The first use of the scale showed a very definite gain over the gradings without the scale. The average deviation from the true score for the first use of the scale was 5.6. This is more than 40 per cent lower than the deviations without the use of the scale. This indicated that even those without practice in the use of a scale improve as the result of the use of objective standards for comparisons. Some investigators have not found so much improvement in the first use of a handwriting scale. But some of these studies have been made with teachers familiar with the grading of handwriting. This group of students were without this practice in grading, therefore they may have received greater benefit from the first use of a scale than teachers.

The result of practice in grading about 30 specimens of handwriting and then checking up on these scores by comparing them with the true scores of these samples resulted in a still further reduction in the errors in grading. The deviation in the fifth grading of the 10 samples was 4.0. This was a 30 per cent reduction over the previous grading. The 8 men who devoted a second practice period to the 30 samples with known scores reduced their average deviation to 2.6. This is almost a 20 per cent reduction for these graders as the result of the extra hour of practice.

This experiment is not essentially different from the method suggested by Thorndike in the monograph already cited. It is interesting in that it shows the large errors in the grading of students without the use of a scale and the rapid improvement in grading as a result of a small amount of practice in grading samples of handwriting with known score values. The improvement in this experiment is much

more rapid than found by C. T. Gray.[1] In Gray's experiment there were no true values for the specimens by which the graders could check up on their scoring. The improvement came only as a result of practice and conference following the grading. Even here there was improvement but it was not nearly so rapid as that found in the above experiment. It is to be hoped that a list of samples similar to the Thorndike set will be devised for use with the more widely used Ayres handwriting scale. The author hopes to prepare such a list with true values in terms of the Ayres scale. By a relatively short training period—even an hour—the average teacher without experience in the use of a handwriting scale should be able to reduce her errors in grading to 4 per cent or less.

Summary

1. The results of this study show a very slight improvement in regrading samples of handwriting by the traditional methods. This improvement is practically negligible and the amount of error and the variation from one grading to another indicates the unreliability of the traditional methods of subjective grading.

2. In this experiment there was a decided improvement in grading as a result of the use of a scale as compared with the usual subjective method of grading handwriting.

3. The effect of one or two hours' practice in grading samples of handwriting by the use of a scale followed by a comparison of the grades given with the true values of these samples resulted in a much greater improvement in the grading. As a result of such practice most of the students of this experiment were able to give grades to samples of handwriting within 2½ to 4 per cent (AD) of their true value. Such grades are unusually reliable and the method here used suggests a means for improving the grading of handwriting.

[1] C. T. Gray.: The Training of Judgment in the Use of the Ayres Scale for Handwriting. *Journal of Educational Psychology*, Vol. 15, 1915, pp. 269–78.

A SCORING STENCIL FOR MIRROR-DRAWING

HAROLD E. ISRAEL

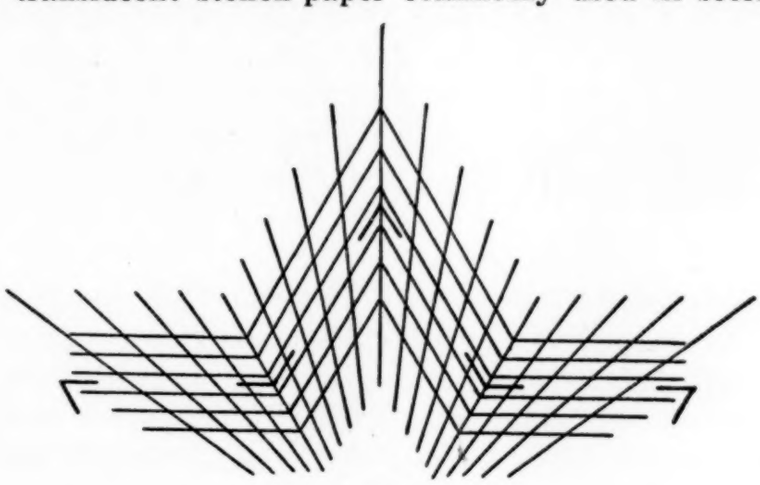
Dartmouth College

Although most teachers of psychology make some use of mirror-drawing, either as demonstration, experiment, or diagnostic test, there seems to be little accuracy and no uniformity in the methods of scoring errors. It is seldom practicable to have an observer record the errors on a counter (as suggested by Whipple), especially if a long series of tracings is to be made over a period of time. The stencil partially reproduced in the figure below provides a quick and uniform method for counting errors in the star tracings, making scoring possible at any convenient time after the tracings have been completed.

The stencil is fitted over the star by means of the short guide lines at the points; then the traced outline is followed around from the starting point. Whenever any one of the boundaries of the zones parallel to the outline is crossed, an error is counted (naturally recrossings in the corrective direction are not counted). Thus a deviation beyond the extreme limits of the stencil in either direction from the outline will give an error score of three. Moreover, errors are added for each radial boundary, over one, which the tracing crosses before returning to within the "home" parallel-zone. Allowing thus the width of two radial zones for the correction of a wrong motion, this feature penalizes hasty cutting across or carelessness.

A SCORING STENCIL FOR MIRROR-DRAWING

As used in our laboratory these stencils are drawn in India ink upon the translucent stencil-paper commonly used in scoring tests.



A much more accurate and permanent form is made by scratching the outlines upon transparent celluloid and then staining with ink. The figure on p. 537 is drawn to scale from the star blank probably in most general use (Stoelting, No. 31,205). Lateral (parallel) zones are 3 mm. in width, and radial zones are 6 degrees. The same scheme can readily be applied to the double-outline blanks prescribed by Strong and Foster, although the size of the printed blanks in current use is slightly different from that of the single-outline star to which the figure is fitted.

CORRELATIONS IN PHYSICAL AND MENTAL GROWTH (*Concluded*)

ETHEL M. ABERNETHY

Queens College

PART II

ANALYTICAL STUDY OF CASES EXCEPTIONAL IN PHYSIOLOGICAL DEVELOPMENT

The problem of relationship between physical and mental growth was attacked from another angle through an analytical study of cases

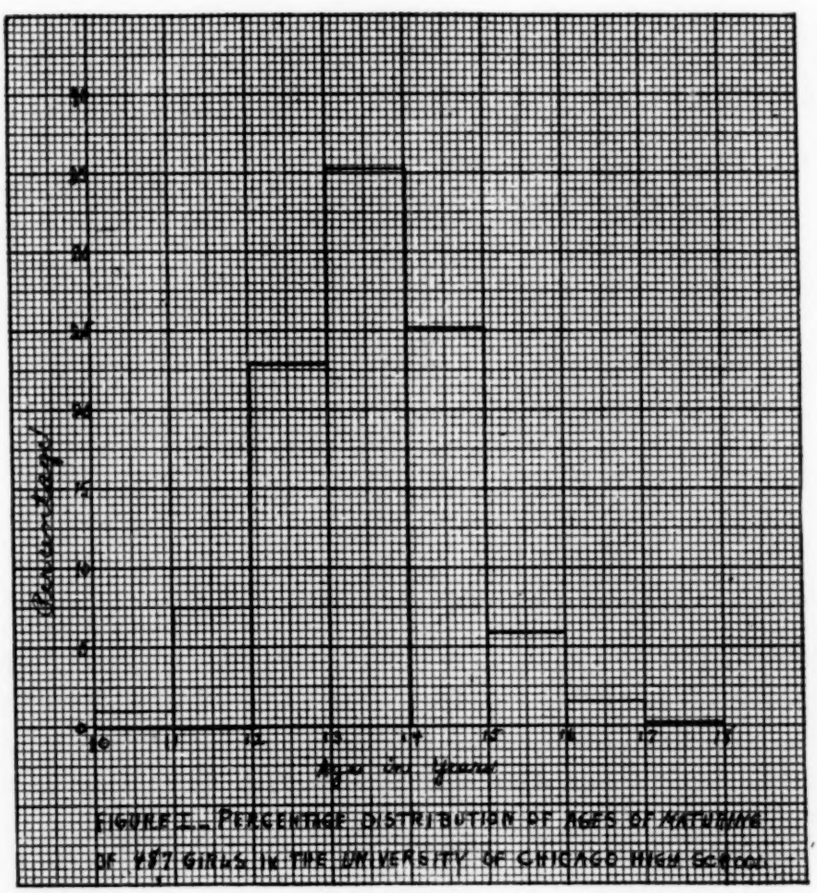


FIG. 1.

exceptional in age of physiological maturation as indicated by pubescence. If any correlation exists between intelligence and physical status, it seems reasonable to expect it to appear in cases markedly different from the average in physical growth. Pubescence is probably the most commonly accepted standard of general physical matur-

ing. Study was therefore made of individuals noticeably above or below the average age of major adolescent change.

In order to determine the average chronological age of pubescence for girls in the University of Chicago High School, examination was made of the physical records of 487 girls recently enrolled within this school. A distribution of these ages of maturing gave a nearly normal curve, with the average age, by any measure of central tendency, approximately 13 years, 6 months. Baldwin found the median age of maturing for 56 girls in the same school to be 13 years, 9 months, and concluded that Chicago girls mature later than girls of the country and smaller city.[1] Distribution of a much larger number of cases shows that the average age of maturing for these girls is 3 months earlier than Baldwin's data led him to believe. The data from the present investigation are presented in Table X and in Fig. 1.

TABLE X.—PERCENTAGE DISTRIBUTION OF AGES OF MATURING OF 487 GIRLS IN THE UNIVERSITY OF CHICAGO HIGH SCHOOL

Age	Number of cases	Per cent
10	5	1.03
11	37	7.60
12	112	22.99
13	172	35.32
14	123	25.26
15	29	5.96
16	8	1.63
17	1	.21

This distribution shows agreement with the general finding that range in age of maturing is wide. For more than a third, however, pubescence occurs between the thirteenth and fourteenth birthdays.

For this study there were selected 38 girls who in age of maturing differ widely from the average. Nineteen are accelerated in physiological maturing; 19 are retarded. Those accelerated were pubescent between 10½ and 11½ years of age. For those retarded pubescence occurred—in 16 of the 19 cases—between ages 15½ and 16. Three of the girls retarded in development are nearly 17 and are not yet mature. A comparison of the two groups will first be given, to be followed by

[1] Baldwin, B. T.: "The Physical Growth of Children from Birth to Maturity." p. 191.

a descriptive study of a few typical individuals from the retarded group.

Group Comparisons

In studying the accelerated and retarded in maturing, attention was given to intelligence test scores, grade location, class standing, social development, general physical growth, health, and adjustment to the demands of the school. These phases of the comparative study are summarized in order of mention.

1. *Intelligence Test Scores.*—The statistical study previously reported gave no evidence of a correlation between precocity of maturing and mental ability as measured by an intelligence test. When the extremes in age of pubescence are compared there is still no significant relationship. The median IQ for the retarded is 112; for the accelerated, 114. Moreover, there is hardly any difference in range of intelligence quotients for the two groups.

2. *Grade Location.*—Another measure of intellectual capacity may be found in progress through the grades. The following table presents a comparison of the two groups with respect to grade location.

Table XI.—School Progress of 19 Pupils Retarded and 19 Pupils Accelerated in Physiological Maturing

School progress	Physiologically retarded	Physiologically accelerated
Accelerated two years	..	2
Accelerated one year	1	9
Normal	15	6
Retarded one year	1	1
Retarded two years	2	1

Of the 19 accelerated in age of maturing, 11 are above the grade normal for their age; while only 1 of the group physiologically retarded is above the average in school progress. In grade location the median pupil of the physiologically accelerated group is one year ahead of the physiologically retarded group.

3. *Class Standing.*—It is important to know not only the grade location of a pupil but that pupil's success within the grade. Table XII shows teachers' ratings of the pupils in these two groups.

TABLE XII.—TEACHERS' RATINGS OF 19 PUPILS RETARDED AND 19 PUPILS ACCELERATED IN PHYSIOLOGICAL MATURING

Teachers' ratings	Physiologically retarded	Physiologically accelerated
Superior	2	2
Above average	2	5
Average	6	8
Below average	6	3
Inferior	3	1

Comparison of these ratings indicates that those who matured early are somewhat more successful even in the more advanced work attempted.

4. *Social Development.*—Many types of sources were consulted in order to gain the idea of the general social development of each of the girls studied. Of the 38 pupils in the two groups, 10 are rated by teachers and deans as noticeably deficient in such social qualities as cooperativeness, leadership, consideration of others, and assumption of responsibility. Nine of the 10 belong to the group of the physiologically retarded. For these 38 girls there is a pronounced relationship between physiological age and social age.

5. *General Physical Growth.*—In general, early adolescence seems to be more favorable to physical growth than very late adolescence. For a few individuals pubescence after the fifteenth year may be normal, but for these cases of extreme retardation 16 of the 19 are below the norm for their age in height and weight, in many cases very markedly below average. Although the correlation for larger groups failed to show any influence of age of pubescence upon ultimate size, it seems highly probable that in extreme cases of retardation the normal development is not usually attained.

6. *Health.*—The cards kept by the school physician give each pupil's health record from the time of enrollment in the school. Of the 19 girls slow in maturing, 11 have been reported as being in poor physical condition for a period covering a year or more. Extreme nervousness and malnutrition are the common ailments. For the 19 who matured about the age of 11 there is only 1 record of ill health.

7. *Adjustment to the Demands of the School.*—There are more "problem cases" among those who matured late than among those who matured early. In some cases difficulty in adaptation to school requirements is due to ill health, in others to social immaturity. How-

ever, a few girls for whom pubescence has been late have been very satisfactory students. Late maturing seems in three cases to be normal from every point of view, probably because of certain racial characteristics.

Individuals Retarded in Maturing

Typical individuals of the physiologically retarded group have been selected for a descriptive account. Each pupil's index number in the University High School will be substituted for her name.

Pupil 614 is a Jewish girl 17 years of age whose general physical development, except for pubescence close to her sixteenth birthday, has conformed to that of the average girl of her race. For two years she has been under treatment for inactive thyroid, but until the age of 15 her health was excellent and rate of growth seemed normal.

In mental ability, as measured by the intelligence test, Pupil 614 is above the general norm. Her IQ on the Otis Higher Examination has averaged 109, but in grade location she is one year retarded while her work is below the average of her class. Her teachers report that she has ability but lacks ambition and initiative. In her study habits she is inclined to be somewhat indolent and careless. Recently, however, there has been improvement in her work.

Adjustment to the social and moral standards of her group was for a long period rather difficult for this girl. For two years she was a serious problem case in the school, although she had been trained in the highest type of Jewish home. Social and moral difficulties, together with inferior class work, made her a very undesirable member of the group. This year, however, there is evidence of a better attitude toward work and group standards. The change has been gradual but apparent.

Pupil 8 was nearly 16 before she was physiologically mature. In general physical development she is below the average. On her seventeenth birthday she measured in height 61 inches (norm 63.6), in weight 99 pounds (norm 121.8), and in ossification ratio 1.04 (norm 1.096). For more than a year before her maturing nervousness and overanxiety about her studies made work so difficult that it was found necessary to arrange a very light school program for her. After pubescence health improved and full work was resumed.

In intellectual traits and scholarship, Pupil 8 has had high rank from the beginning of her high-school career. Before pubescence,

however, there was frequent mention in school reports of a certain mental and social immaturity surprising in one of so bright a mind. Nearly a year after the girl's maturing one of her teachers added to her report, "R. has improved wonderfully since two years ago. What has wrought the change?"

Slow physical growth is accompanied by slow, or inferior, mental development in the case of Pupil 973. This girl at 15 years of age has not yet matured physiologically, has a mental age of 13 years, 10 months (IQ 92), and is two years retarded in school progress. Records obtained from the office of the school physician show that there has been slow physical development from infancy. At birth 973 weighed only three and one-half pounds. She grew very slowly and did not walk until she was two years old. After the age of two her health was better and physical development seemed somewhat more normal. Since entering the subfreshman class in high school, there have been reports of nervousness and general ill health. The decided increase in height and weight during the past year—from 58.6 to 60.8 inches in height, and from 75 to 86.5 pounds in weight—probably indicates early approach to physiological maturity. Ossification ratio has now reached 1.11, which is above the norm (1.096) for age 17.

Pupil 973 is retarded socially and scholastically. She seems to put forth effort to meet requirements, but is too weak for the demands made upon her. The following statements from two of her teachers are typical of all reports given: "Her work is weak. She gives good attention and appears to put forth effort, but fails to get results." "She tries hard, but does her work very poorly and only comes up to the standard with great effort." She is further characterized as "a socially very immature little girl."

For pupil 925, growth seems to have been slow in nearly all physical traits. At the age of 15 she is just pubescent and is below the average in height, weight and ossification ratio. The Otis IQ of 112 indicates more than normal intelligence, but this ability has been manifested rather unevenly in class work. Some teachers report that 925 is "Very, very stupid, has poor memory and unusually slow reaction to explanations," while others have found her a good student and careful worker. In general, her work is below the standard of the school and re-teaching is necessary before credit can be given. Pressure alone has kept her normal in grade location.

Social immaturity and lack of intellectual purpose and independence seem to be the characteristics of Pupil 925. The dean of

girls says of her that, in general social attitude, she is hardly more than 11 years of age.

For Pupil 240, pubescence shortly before the sixteenth birthday was accompanied by unusual growth in height and weight for this age. Within a year height increased nearly two inches and weight about 21 pounds. No corresponding spurt in intelligence was noted, as measured by the mental tests, but improvement in social attitudes and mental organization seemed to follow the advent of adolescence. Of unquestioned superiority in general mental ability, with an IQ of 118, this student was for two years a problem to her teachers because of lack of muscular and mental control. Coordination of any sort seemed difficult for her. Teachers reported her to be a "disorganized individual, erratic, undeveloped, lacking in concentration and dependability." At times, however, she did work of a very high quality.

The most noticeable improvement since physiological maturation has been in general coordination and in social traits.

Conclusion

The number of cases studied is too small for generalization, but there is evidence that age of physiological maturing has direct bearing upon social age, school progress, general physical growth, health, and capacity for adjustment to the demands of the school. In this investigation, little or no correlation has been found between mental age and physiological age. However, an intensive study of 38 cases exceptional in age of pubescence leads to the belief that there may be certain mental and emotional traits, not measurable by intelligence tests, which are influenced by physiological development. Of the 19 girls retarded in maturing, 7 were characterized in school reports before their pubescence as "bright enough but mentally immature," or as "childish and immature in attitude." For 3 of the girls studied, late maturing was apparently normal, and due, probably, to racial or family heredity. In general, however, pubescence after the fifteenth birthday was found to be less favorable than very early maturing.

General Summary

1. Statistical study of the physical and mental measurements of 359 girls reveals little or no relationship between mental age and the generally accepted criteria of physiological and anatomical development. When chronological age is held constant, zero or slightly negative coefficients are found in the correlating of mental age with

the ossification ratio, with number of permanent teeth, and with precocity in physiological maturation. Weight and height have only zero or low positive correlation with mental age.

2. In general, coefficients are also low in correlations between physical traits. The ossification ratio has a low but positive correlation with height and with dentition. There is marked relationship, however, between carpal development and precocity in pubescence.

3. Except in cases of extremely late maturing, there is no evidence that ultimate size is influenced by the age of pubescence.

4. For 38 girls exceptional in age of maturing there is evidence that physiological age is directly related to general physical and social development and to school progress. Although intelligence test results indicate no relationship between mental ability and age of pubescence, it seems probable that physiological maturing has bearing upon important intellectual, emotional and volitional traits which are not measurable by intelligence tests.

5. Pubescence after the fifteenth birthday is, in the majority of the cases studied, accompanied by ill health and inability to meet the requirements of the school. Such deviation from the average seems, except for a few individuals, to be abnormal. Very early pubescence is more favorable to general development.

6. The number of cases studied does not justify final conclusions. There is suggestion for further investigation in the relation of age of maturing to general school progress and in the correlation of precocity of maturing with the ossification ratio.

THE MENTAL ATTAINMENTS OF COLLEGE STUDENTS IN RELATION TO PREVIOUS TRAINING[1]

JOHN W. GOWEN AND MARJORIE GOOCH

These papers present an analysis of those combined factors of intelligence and training which make for mental success in college students. Beginning with the ideal of training for culture, our present day educational system has developed through successive stages until the present broad, humanitarian ideal of training for citizenship has come to be generally accepted. It is but yesterday that many of our colleges dropped their classical requirements for all entrants, be they planning for engineers or statesmen. The evolution has been slow, but it is just as real as that found in other sciences. During this period of growth and change, and particularly during the last 20 years there has accumulated a vast body of unanalyzed data pertaining to education and those related mental processes which expedite it. These data, while not always accurate, are found on the books of almost every college and university. It is the purpose of this investigation to study these records in one institution, the University of Maine, in the hope of throwing some light on certain of the problems open to analysis by the use of such materials. The relation of mental attainments in such high school subjects as English, chemistry, algebra, plane geometry, Latin, elementary and advanced French with such freshmen year college subjects as English, chemistry, algebra, analytical geometry, elementary and advanced French, and elementary German, is one such problem. Other variables such as sex, age, and quality of previous training as measured by the high school will be taken up in later papers.

METHODS

The records from which our data were obtained are those of the Registrar of the University. Of the many facts contained in the college records, we were interested more particularly in the grades received in those subjects which a preliminary survey had shown to be most frequent. These subjects are in order of their occurrence, English, chemistry, mathematics, French, German, mechanical drawing, biology, history, physics, and Spanish. The subjects and order are essentially the same in college and high school.

[1] University of Maine Studies.

These data were subject to the following limitations: in order that the ranks might be as comparable as possible only such students as entered between 1909[1] to 1917, inclusive, and who took a straight course, receiving a bachelor's degree, were used. No transfers from other colleges, law students, school course agricultural students, nor specials were considered. The college grades used were limited to those of the freshman year, since the work that year would seem to be most closely related to the high school work. With these restrictions there are 927 graduates from the University of Maine who have records for the freshman year and high school.

Because of the small numbers in many courses, we have limited ourselves to consideration of the following subjects. In college: English, chemistry, algebra, analytical geometry, elementary French, advanced French, and German. In high school: English, chemistry, algebra, plane geometry, Latin, elementary French, and advanced French.

The following numerical values have been assigned: The high school A = 90 to 100, with the mid-point of the interval 95; B = 80–90, mid-point 85; C = 70–80, mid-point 75; D = 60–70, mid-point 65; E = 50–60, mid-point 55. The college A = 95–100, mid-point 97.5; B = 85–95, mid-point 90; C = 75–85, mid-point 80; D = 65–75, mid-point 70; E = 55–65, mid-point 60; F = anything below 55, regarded for convenience as having a mid-point of 50.

Having fixed the limits to the data, correlation tables were made to show the mental attainments of the student in high school as contrasted to his attainments in college. The ordinary correlation coefficient of Pearson was used to measure the degree of relationship. The tables are deposited in the Library of the University of Maine.

Several problems arise in the consideration of this material. The differences in the quality of the high schools offer a possible cause of variation. Approximately 300 high schools are represented by the data used. These schools differ widely in standards of scholarship, courses, and grading systems. The standard of scholarship might allow students to graduate from one high school, that a second high school would not consider worthy of graduation. Differing courses in high schools may result in lack of adequate training in actual subject-matter. For even the most brilliant student would be expected to find himself at a disadvantage if the rest of the class has had much

[1] In the fall of 1909 the present system of grading (A, B, C, D, E, F) was substituted for the somewhat more complicated system in use previously.

better fundamental training. A great deal of the work taken the freshman year is directly dependent in subject-matter on a preparatory course. A poor preparatory English course even though the student did rank well in it might handicap a freshman so that his college grade would be lowered below what his real ability would warrant. The differences in grades in high school may be of two sorts. There are some differences that are due to the difference of values attached to the letters used and to the use of a percentage system. As far as possible these were corrected for when the cards were punched. But the second sort of differences in grades, those where the personal equation of the teacher was involved, is an important problem, an analysis of which will be presented later.

In continuing an investigation having to do with college courses back even a few years, we are at once brought face to face with the fact that college courses change. Usually the course number, and to a great extent the course content, varies with the department's head. When the latter changes, the former are apt to. This is a real source of error. Yet as far as possible, this has been guarded against by the grouping together of only such courses as seemed equivalent according to the college catalog's description and the experience of the writers. This hypothetical difficulty may be finally tabled by saying that investigation showed that the average rank and the standard deviation of this rank remained quite uniform in all subjects during the time covered by the data used in these researches.

Due to the fact that entrance records are kept on several differing types of certificates, some of which do not record high school ranks, but only subjects in terms of units, our correlations with high school grades are not so complete as might be desired. All the tables, however, are subject to the same limitations.

The records of any group of college students are, of course, based on a selected population. Our conclusions are consequently subject to a number of qualifications as to their applicability, because of both the character of the material and the selected nature of the population. These very restrictions give the conclusions much of their importance, as all the deductions rest on the every day data of the college and high school and are applicable to these data under the variety of conditions which surround them. The problems approached and the answers returned are, consequently, the great practical problems of modern organized society.

The Effect of Previous Training as Measured by College Grades

The effect of previous training on the resulting mental attainments of college students is a subject which offers many avenues of approach. The work of a number of different investigators with different subject-matter and differing nomenclature, seems to lead to the idea that any transfer of training takes place by means of identical elements, either in content or procedure. In view of this hypothesis can it be said that the brighter than average high school student tends to become a brighter than average college student? If so, can this be ascribed to transfer of training? Is there a strong correlation between subjects showing marked similarity of subject-matter or are the correlations between all subjects somewhat uniform? Should the relation between the grades prove to be somewhat uniform it would suggest an identity common to all accumulation of knowledge for any subject taken by the student.

Our conclusions are based on high school ranks as indicators of subsequent work. It is realized that this index is only relative. The work of Thurstone[1] shows clearly that success in high school as indicated by the rank of the student, is predictive of future success in college. Thus he found that the average correlation between high school subjects and first term freshman subjects for engineers in college was .297 whereas the average correlation for a group of six tests designed to determine the individual mental equipment for college was .324. Thus high school ranks were shown to be fairly good indicators of college work, in fact almost as good as a series of tests planned expressly for this.

In general, we do not find a distinctly higher average in college marks than in high school. This does not mean, however, that high school work has no effect on college work. In fact, if the selective scale of marking remains the same the average marks in college and high school should also remain the same even though the college population may represent a more highly selected group. Table I shows the means, standard deviations, and correlation coefficients of the various college subjects with high school English.

The first thing noticeable about the correlations between these subjects is that they are all significant. That is, the r's or correlation

[1] Thurstone, L. L: Intelligence Tests for Engineering Students. *Engineering Education*, Vol. XIII, 1923, No. 5, pp. 263–318.

TABLE I.—RELATION OF GRADES IN COLLEGE SUBJECTS TO HIGH SCHOOL ENGLISH

High school grades				College grades			
Subject	Mean	SD	*r*	Subject	Mean	SD	*n*
English.....	84.6 ± .3	7.95 ± .16	.284 ± .026	English.....	81.5 ± .3	9.12 ± .18	563
English.....	84.2 ± .2	7.86 ± .17	.209 ± .030	Chemistry.....	80.5 ± .3	9.18 ± .20	471
English.....	83.8 ± .3	8.31 ± .21	.220 ± .033	Algebra.....	82.1 ± .4	11.59 ± .29	361
English.....	83.7 ± .3	8.33 ± .22	.276 ± .033	Analytical geometry.	82.2 ± .4	10.34 ± .27	342
English.....	85.3 ± .8	7.16 ± .56	.403 ± .092	Elementary French.	81.6 ± 1.1	10.16 ± .80	37
English.....	86.5 ± .5	7.88 ± .33	.389 ± .050	Advanced French...	86.9 ± .5	9.13 ± .38	129
English.....	84.4 ± .4	8.01 ± .25	.190 ± .041	German.....	82.7 ± .5	9.59 ± .30	240

coefficients are all enough larger than their probable errors to make the chances of their being due to random sampling, small. For instance, the correlation of English and English is only .284 yet it is 11 times its probable error.

In other words the student who got high rank in high school English would tend to get high rank in college English. This is, however, only a tendency. The actual position of a given student would not be the same in high school and college. In fact, the average student who averaged A (97.5) in college English, averaged 88.6 in high school English. Those that averaged B, which is considered 90, averaged 87.6 in high school, those that averaged C or 80, averaged 83.6, those that averaged D or 70, averaged 81.4, those at E or 60 averaged 78.8 in high school, and those at F, averaged 83.9. These last two, however, are not significant, for they were based on only eight and nine individuals. Table II shows these averages.

TABLE II.—AVERAGES IN COLLEGE ENGLISH COMPARED WITH AVERAGES IN HIGH SCHOOL ENGLISH

College grade in English	*n*	High school grade in English
A or 97.5	33	88.6 probably significant
B or 90	167	87.6 significant
C or 80	247	83.6 significant
D or 70	99	81.4 probably significant
E or 60	8	78.8 not significant
F or 50	9	83.9 not significant

Disregarding the last two which are not significant because of the smallness of the populations, we can see that while the values do not correspond very exactly, yet the two series decrease in the same order.

Comparing these values with the correlation between high school English and advanced college French, we find a slightly higher correlation coefficient .389 ± .050. Yet, the probable error of the difference between this correlation and that of English with English is almost as large as the difference, making the difference not significant. Expressed in averages the relation between high school English and college French is shown in Table III.

TABLE III.—AVERAGES IN COLLEGE FRENCH COMPARED WITH HIGH SCHOOL ENGLISH

College French	*n*	High school English
A or 97.5	25	91.8 significant
B or 90	60	87.3 significant
C or 80	30	82.0 significant
D or 70	11	84.1 probably significant
E or 60	2	75.0 not significant
F or 50	1	85.0 not significant

Here again the last two figures are unreliable as they are based on such very small numbers (2 and 1). In fact, the higher probable error of the correlation coefficient is due to the smaller population concerned. The correlation of English and elementary college French is based on so few numbers as to make it unreliable.

The difference between the correlation of English and English and the correlation of English and German, again, is scarcely significant for it is less than twice the probable error of the difference. So that although there appears to be quite a difference (.284 to .190) yet this difference may be attributed to random sampling. The comparison of the averages will show a marked unevenness in the value of the E and F due to lack of numbers A in college German, or 97.5, averages 89.1 in high school English; B averages 86.2; C, 82.6; D, 79.1; E, 91.7; F, 90.0. Table IV gives the summary of these averages.

Some of these differences in accomplishment in German and English are clearly due to small numbers. Generally, however, the person who

TABLE IV.—AVERAGES IN COLLEGE GERMAN AS COMPARED WITH HIGH SCHOOL ENGLISH .

College German	*n*	High school English
A 97.5	29	89.1 significant
B 90	66	86.2 significant
C 80	105	82.6 significant
D 70	29	79.1 significant
E 60	9	91.7 not significant
F 50	2	90.0 not significant

is a good student in high school English will tend to be better than the average in college English, French, and German.

It is commonly supposed that the sciences require a different sort of mental ability than is required by the languages. By considering chemistry, algebra, and analytical geometry as forming a science group and English, elementary and advanced French, and German as a language group, an average weighted correlation coefficient of .234 may be obtained for the science group and .291 for the language group when the correlations are with high school English. The square root of *n* is the basis on which the coefficients are weighted. If language ability is transmitted separately from scientific ability we should expect the correlations with other languages to be high and the correlations with the sciences to be low. While the difference in the weighted correlation is in the right direction, it is so small as to be of only doubtful significance. Such being the case and insofar as English is concerned it does not seem as if this college population can be grouped into "those good in languages," and "those good in sciences." That is, there does not seem to be distinct levels between those subjects

TABLE V.—AVERAGES IN COLLEGE ALGEBRA COMPARED WITH HIGH SCHOOL ENGLISH

College algebra	*n*	High school English
97.5	57	86.4 significant
90	105	85.4 significant
80	111	82.8 significant
70	52	81.4 significant
60	31	81.1 significant
50	5	81.0 not significant

which show a marked similarity of subject-matter and those which do not. The range of averages between English and algebra, for instance, shows that the relations are about the same as those between the languages.

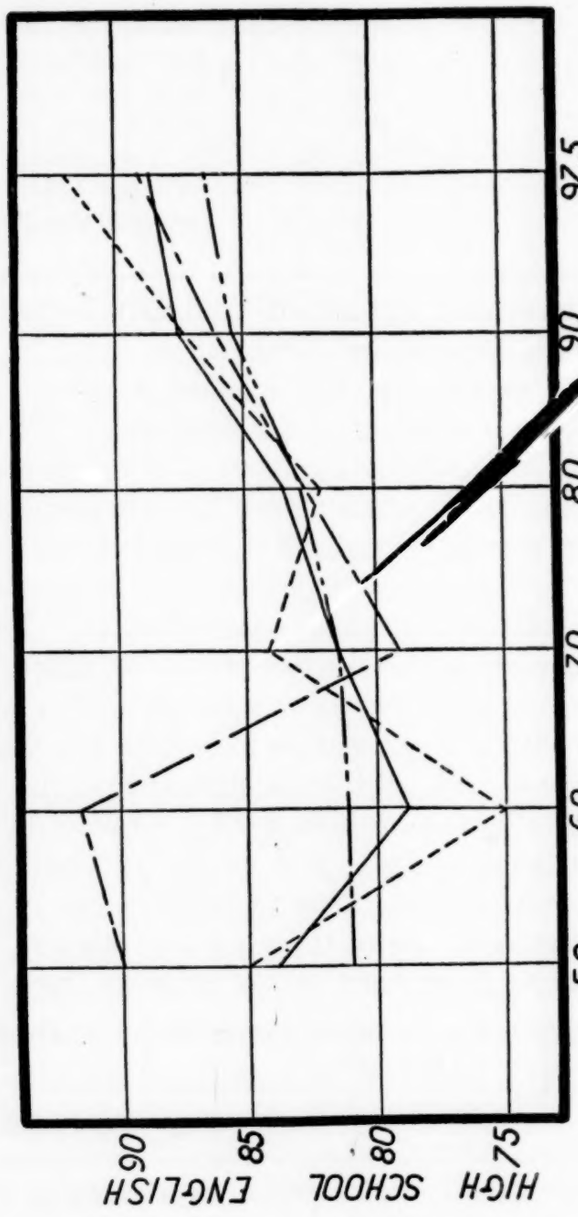


FIG. 1.—College attainments compared with grades in high school English. The solid line represents college English; the broken line college French; the dot and dash line college German; and the dot dot dash line college algebra.

While the actual values of the correlation coefficients for high school English and college French are greater than the correlations of English and the sciences, these differences are not significant because of the smaller number of individuals concerned in the former case.

The comparison of college grades in English, advanced French, elementary German, and algebra with the average grades in high school English is shown in Fig. 1. The graphs show that high class collegiate work is based on a relatively high grade of high school work. This conclusion holds for the principal subjects studied in college and high school.

It can be shown that in the correlation of high school chemistry with the various college subjects, the difference between even the highest coefficient $.335 \pm .036$ and the lowest $.181 \pm .080$ is not significant, being less than twice the probable error of this difference. Here again we get no very clearly defined groups, the weighted average correlation coefficient of high school chemistry with the languages being .216 and with the sciences .276.

The relations of high school chemistry and the college subjects are given in Table VI.

TABLE VI.—RELATION OF MENTAL ATTAINMENTS IN COLLEGE SUBJECTS TO HIGH SCHOOL CHEMISTRY

High school grades				College grades			
Subject	Mean	SD	r	Subject	Mean	SD	n
Chemistry.....	85.0± .3	8.23± .20	.198± .032	English.....	80.7± .3	9.79± .23	405
Chemistry.....	85.3± .3	8.30± .21	.193± .035	Chemistry.....	81.1± .3	8.44± .21	354
Chemistry.....	84.8± .3	8.49± .24	.335± .036	Algebra.....	81.8± .5	11.63± .33	276
Chemistry.....	84.8± .4	8.48± .25	.311± .037	Analytical geometry.	81.7± .4	10.54± .31	263
Chemistry.....	85.8±1.2	8.74± .81	.256± .121	Elementary French..	82.2±1.3	9.57± .90	26
Chemistry.....	85.9± .7	8.05± .47	.181± .080	Advanced French...	85.8± .7	8.98± .52	67
Chemistry.....	85.1± .4	7.88± .21	.250± .046	German.....	82.0± .5	9.20± .32	190

The tables of the average ranks are omitted as they offer no striking differences from those given under English in high school. If we continue the comparison of ability in science and in language it is noticeable that with chemistry in high school the weighted average correlation with sciences in college is larger than with languages, being .276 in the former and .216 in the latter case, while with English in high school the converse is true—the weighted average correlation with languages is higher.

On the hypothesis that ability in the sciences is something distinct from ability in the languages, these differences are both in the right direction. The actual differences are small and in comparison with the probable errors are not significant.

Table VII points out the relations and differences between high school algebra and the same college subjects which have been considered previously.

TABLE VII.—RELATION OF MENTAL ATTAINMENTS IN COLLEGE SUBJECTS TO HIGH SCHOOL ALGEBRA

High school grades				College grades			
Subject	Mean	SD	*r*	Subject	Mean	SD	*n*
Algebra.....	85.0± .2	8.23±.16	.184±.027	English.....	80.4± .3	9.70±.19	604
Algebra.....	84.7± .3	8.62±.19	.211±.030	Chemistry.....	80.7± .3	9.19±.20	463
Algebra.....	84.8± .3	8.76±.22	.407±.031	Algebra.....	81.7± .4	11.95±.33	352
Algebra.....	84.8± .3	8.70±.23	.384±.032	Analytical geometry	82.1± .4	10.66±.28	330
Algebra.....	83.4±1.0	9.16±.72	.427±.091	Elementary French..	81.6±1.1	10.16±.80	37
Algebra.....	86.6± .5	7.47±.32	.259±.055	Advanced French...	86.4± .6	9.24±.40	122
Algebra.....	84.9± .4	8.50±.27	.313±.041	German.....	82.5± .4	9.61±.30	230

Table VII shows a significant difference between the correlation of algebra in high school with algebra in college, on the one hand, and the correlation of algebra in high school with English in college, on the other. The weighted average for science is higher, but not significantly higher, than the weighted average of the languages. These averages are .329 for science against .258 for language.

Table VIII shows that geometry in high school has a higher correlation with the mathematics courses in college than with some of the other courses. The correlation between geometry and analytical geometry (.337±.032) is almost four times the probable error of the difference greater than the coefficient for geometry and English (.190±.028). If we get the weighted average of these correlations, however, we find the science group has an average correlation of .321 while that of language is .176. These correlations are on the borderline of significance.

The correlations with Latin in high school and the college subjects as given in Table IX bear out the general theory that there is not a marked difference between those good in science and those good in languages. The weighted average correlation is slightly higher for the language group, (.281) than for the science (.261).

It is commonly argued that Latin furnishes an excellent preparation for French because of the resemblance between the languages. The resemblance between Latin and German is not generally considered so marked. The evidence here presented, however, shows that the

TABLE VIII.—RELATION OF MENTAL ATTAINMENTS IN COLLEGE SUBJECTS TO HIGH SCHOOL GEOMETRY

High school grades				College grades			
Subject	Mean	SD	r	Subject	Mean	SD	n
Geometry.....	84.8± .3	9.08±.19	.190±.028	English.....	81.2± .3	9.74±.20	548
Geometry.....	85.1± .3	8.83±.20	.227±.031	Chemistry.....	80.7± .3	9.19±.21	449
Geometry.....	85.3± .3	9.32±.24	.409±.030	Algebra.....	82.0± .4	11.73±.30	344
Geometry.....	85.5± .3	9.00±.24	.337±.032	Analytical geometry.	82.0± .4	10.78±.29	325
Geometry.....	84.4±1.0	8.72±.71	.135±.110	Elementary French..	82.0±1.2	10.39±.85	34
Geometry.....	84.9± .5	8.77±.37	.031±.060	Advanced French...	86.7± .6	9.32±.40	126
Geometry.....	85.1± .4	8.83±.28	.279±.041	German.....	82.5± .4	8.34±.27	223

students good in Latin were slightly better in German than they were in French. The supposed similarity between Latin and French does not seem to have influenced the students markedly.

This evidence does not entirely settle the problem, however. The advantage of Latin may be due to have taken it at all *vs.* not having taken it instead of the grade actually obtained in the Latin courses. The differences derived from this comparison do not particularly favor Latin as a training subject for French, however. Thus advanced French students taking Latin in high school received an average rank of $88\pm.6$ whereas those not taking Latin received a rank of $83\pm.9$. With German the differences are in the same direction. The German student taking Latin in high school had a rank of $83\pm.5$ and those not taking Latin in high school had $80\pm.6$

TABLE IX.—RELATION OF MENTAL ATTAINMENTS IN COLLEGE SUBJECTS TO HIGH SCHOOL LATIN

High school grades				College grades			
Subject	Mean	SD	r	Subject	Mean	SD	n
Latin.....	84.3± .3	8.57±.20	.229±.031	English.....	81.7±.3	8.84±.21	424
Latin.....	84.2± .3	8.44±.22	.228±.036	Chemistry.....	80.8±.3	9.07±.24	331
Latin.....	84.1± .4	8.56±.25	.259±.039	Algebra.....	82.8±.5	11.36±.34	255
Latin.....	84.0± .4	8.49±.26	.297±.040	Analytical geometry.	82.4±.5	10.79±.33	242
Latin.....	85.4±1.1	8.38±.77	.137±.126	Elementary French..	85.9±.9	7.22±.66	27
Latin.....	86.5± .6	8.41±.40	.316±.060	Advanced French...	87.5±.6	8.93±.42	103
Latin.....	83.8± .4	8.12±.29	.387±.042	German.....	83.4±.5	9.43±.33	181

Attention may be called to the fact that the relations between the grades in Latin with those of other subjects are, in general, no higher than those shown by the other subjects compared. Such being the case it would appear that Latin had no better training value than had a more modern subject which also has the advantage of a practical use in our present day life.

A curious thing is shown in Table X. The correlation between elementary French in high school and advanced French in college is almost identical with the correlation between elementary French and algebra, .411 and .408, respectively. The correlation with German is .373 and with analytical geometry .375. Weighting these correlations, no significant difference is found between the language group and the science group, the former being .326, the latter .345. Not only is there no significant difference but what difference there is, is in the wrong direction for if identity of subject-matter were a distinct factor influencing the transfer of ability we would expect a distinct correlation between elementary French and advanced French and one significantly higher than that between elementary French and algebra.

In considering the ranks in advanced French in high school with corresponding ranks in other subjects in college, it is again noticeable that the correlations are higher than those in the Tables with high school English, for example. This is illustrated by Table XI. These correlation coefficients are also more uniform than those found for the other high school subjects.

The weighted average of the science group is .323. The weighted average for the language group is .355. The difference between these groups is small and not significant if judged by its probable error. The difference even though in the right direction is not sufficient to lend any support to the theory of the individual having ability in like subjects and not having the same degree of ability in unlike subjects.

These data bear on two major problems: first, the problem of whether, in any one individual, there is a level of mentality, which is an attribute of that individual, relatively consistent under like conditions; second, whether mentality itself is a composite made up of specific ability in certain lines of endeavor, or a whole largely affecting all courses of study equally, irrespective of the subject-matter which they may contain. The data, herein presented, furnish clear, straightforward evidence for the view that when a student reaches college he is equipped with a mental ability which is a true attribute, capable of finding expression in his college work in accord with the expression

TABLE X.—RELATION OF MENTAL ATTAINMENTS IN COLLEGE SUBJECTS TO HIGH SCHOOL ELEMENTARY FRENCH

High school grades				College grades			
Subject	Mean	SD	r	Subject	Mean	SD	n
Elementary French	83.2 ± .3	8.70 ± .19	.252 ± .028	English	81.2 ± .3	9.79 ± .21	496
Elementary French	82.9 ± .3	8.65 ± .21	.283 ± .031	Chemistry	80.7 ± .3	9.27 ± .22	405
Elementary French	82.5 ± .3	9.04 ± .25	.408 ± .031	Algebra	82.0 ± .5	11.78 ± .32	309
Elementary French	82.3 ± .4	8.99 ± .25	.375 ± .033	Analytical geometry	82.4 ± .4	10.40 ± .32	288
Elementary French	84.4 ± .5	8.78 ± .38	.411 ± .051	Advanced French	86.5 ± .6	9.19 ± .40	120
Elementary French	82.5 ± .4	9.05 ± .29	.373 ± .039	German	82.8 ± .4	9.81 ± .32	219

TABLE XI.—RELATION OF MENTAL ATTAINMENTS IN COLLEGE SUBJECTS TO HIGH SCHOOL ADVANCED FRENCH

High school grades				College grades			
Subject	Mean	SD	r	Subject	Mean	SD	n
Advanced French	84.2 ± .4	9.21 ± .30	.312 ± .036	English	81.6 ± .4	9.59 ± .31	270
Advanced French	84.0 ± .4	9.14 ± .30	.257 ± .041	Chemistry	80.7 ± .4	9.01 ± .29	214
Advanced French	83.9 ± .5	9.47 ± .36	.329 ± .046	Algebra	83.0 ± .7	12.07 ± .46	155
Advanced French	83.5 ± .5	9.43 ± .38	.398 ± .053	Analytical geometry	82.3 ± .6	10.43 ± .41	144
Advanced French	86.8 ± .6	7.89 ± .41	.350 ± .064	Advanced French	88.7 ± .5	6.91 ± .36	84
Advanced French	83.7 ± .6	9.27 ± .43	.434 ± .053	German	83.2 ± .7	9.95 ± .46	105

of it shown by his high school work. Our correlation coefficients for the relation of high school and college work are all markedly significant. They are not so large as some of those presented by certain of the investigators who have studied the subject through the agency of the mental test. The reason for this difference in size is not far to seek since the measures of high school and college ability are subject to many more factors than is a measure on the basis of a mental test. Our data have the advantage, however, of answering the problem as it presents itself to education. For under any system of education more variation is bound to be brought in by the changes of teachers for instance, than by more or less isolated mental tests.

High school and college records represent the success of a student in mental attainments. It is the combined influence of relative intelligence and training which we are studying. These data have the advantage of answering the problem as it presents itself to the educator. While education viewed from the eyes of the younger generation appears to have made marked progress over that of the past, data will be presented to show that there is still much room for further improvement. The limiting factor of this improvement could provisionally be taken as the intelligence of the individual. The weight of present day science shows that this intelligence is due almost entirely to heredity. Consideration of these facts shows the distance of present day education from that of an ideal system making use of all the native ability the student possesses.

It is, then, a practical conclusion that ability in high school indicates to a certain degree future ability in college but that in general this high school ability controls only about 4 to 20 per cent of the variation in college ranks. Thus a poor student in a certain subject in high school may fairly frequently become a rather good student in the subject in college, and a good student in high school may be the reverse in college due to the influence of other variables than the attribute, individual mentality.

The answer to the second problem is equally clear although it is in apparent conflict with what might be expected on the commonly accepted belief that specialized ability is inherited and not generalized ability. This conflict is, we believe, more apparent than real. Our data show that a student who does well in one subject in high school is likely to do well not only in that subject in college but in all his subjects. In other words, there does not appear to be any difference in mental ability in one subject *vs.* ability in the others. Pride

or some such generalized factor appears to be more responsible for the relative level of the rank than any specific factor. The evidence for the belief for specialized ability as an attribute of the individual has grown up chiefly from studies on men and women of great families like those of Darwin and Galton. When these families are compared with our data several striking differences appear. The school of life in which Charles Darwin and Francis Galton won fame accepts only the most brilliant, outstanding work. Of necessity such work has, in general, to be confined to one line of endeavor. The competition brings out in full display only that line of work for which the individual is most fitted, either by environment or heredity, and to which he is likely to contribute most. The competition is a real test of the merit of all the characters of the individual, thus showing in prominence those which are most specific. On the other hand our school system only rarely furnishes such a test. The demand is that the student shall do well, or at least do passing work, in many different lines of study. The student who does one subject easily lays it aside as soon as possible to spend more time on the subject which is harder to him. Our school system tends rather to emphasize a more generalized factor of pride in accomplishing a more equalized grade of work in many lines. This is the factor, whose workings are seen in our study.

Such a view in no way conflicts with the growing views of heredity in accounting for the development and transmission of characters like mental ability, when their expression is estimated in the two different measures of attainment like those of college and those of life. Size as measured by body weight, sex, and life itself have all been shown to be the expression of and to be influenced by many hereditary factors. Evidence is accumulating for the belief that different composites of these factors may give the same end reaction as expressed by the individual. It seems probable, that the brilliant individual in one line of study is likely to have a basic intelligence of a high grade showing in all lines of work, and that the very factors which contribute toward brilliance may be important in the generalized intelligence in the same way that size inheritance is a factor in milk yield in dairy cattle. Under these circumstances, with the situation in our schools favoring generalized ability rather than specialized ability it is clear that the results of this study are in accord with, rather than opposed, to the inheritance studies on specialized traits as exemplified in the studies of the families of the great.

The Presence and Absence Hypothesis as Applied to High School and College Training

The emerging of education made possible by the advances of science from the chrysalis of religious pomp precipitated a conflict between the conservative elements clinging to the standards of the ancient languages and the insurgent leaders who hold that modern French and German are equally good training grounds for developing mental ability. A conflict marked by acrimony and lack of data is scarcely adequate to settle this problem. The rapid trend toward specialization is pressing the productive effort in the most vigorous period of manhood into the senescent years. The ungoverned accumulation of trash littering the path of educational progress in almost any field is inexcusable in this age as the oblivion of many to the appalling proximity of the age when permanent agriculture will no longer support the increasing population. The current capacity of the mind is limited as that of the soil. It cannot produce early if it is fertilized by inert material. The elimination of the unnecessary and the organization of the remainder, offer a bright prospect for educational progress. These thoughts have probably activated the radicals of the educational field.

With the idea of approaching this problem critically, the data used in the previous study have been examined for the influence of training in the subjects of Latin, chemistry, and French. These subjects were chosen for analysis because the two groups, those that had the subject in high school and those that did not, were sufficiently large in numbers to make such analysis worth while. Table XII gives the results from a comparison of the college ranks of students who have studied Latin, French, and chemistry in high school *vs.* those who have not.

The results of Table XII are more easily interpreted by reference to the graphs in Fig. 2. The solid line represents the college ranks of those taking the given high school subject. The dotted line represents the college ranks of those not taking the given high school subjects.

Casual examination of these graphs would lead to the conclusion that the studying of Latin and French in high school was advantageous in increasing the student's mental ability in college but that the studying of chemistry in high school was a disadvantage to the student when he came to college work. Casual conclusions based on almost any experiment involving numbers are generally wrong and so in this

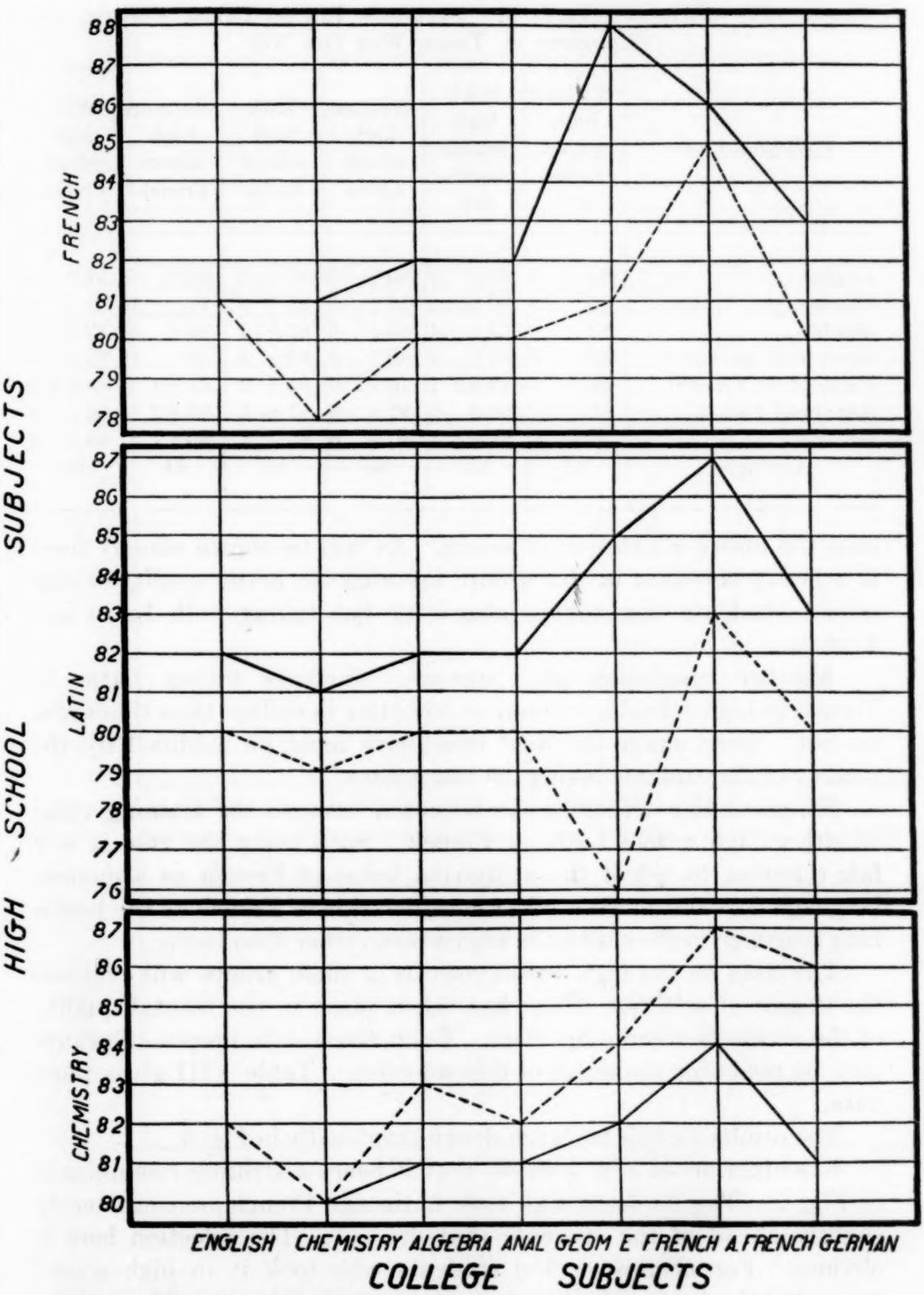


FIG. 2.—Graphs showing the college work of those students who studied chemistry, Latin, or French in high school *vs.* those who did not. The solid line represents those who took the subject; the dotted line those who did not.

TABLE XII.—COLLEGE GRADES FOR STUDENTS TAKING LATIN, FRENCH, AND CHEMISTRY *vs.* THOSE WHO DID NOT

College subject	Without high school chemistry	With high school chemistry	Without high school Latin	With high school Latin	Without high school French	With high school French
English	82 ± .4	80 ± .3	80 ± .5	82 ± .3	81 ± .5	81 ± .2
Chemistry	80 ± .5	80 ± .3	79 ± .5	81 ± .3	78 ± .6	81 ± .3
Algebra	83 ± .6	81 ± .4	80 ± .6	82 ± .4	80 ± .8	82 ± .4
Analytical geometry	82 ± .7	81 ± .4	80 ± .6	82 ± .4	80 ± .9	82 ± .3
Elementary French	84 ± 1.5	82 ± 1.1	76 ± 2.0	85 ± .9	81 ± 1.1	88 ± 1.1
Advanced French	87 ± .7	84 ± .6	83 ± .9	87 ± .5	85 ± 2.8	86 ± .4
German	86 ± .6	81 ± .4	80 ± .6	83 ± .4	80 ± 1.3	83 ± .4
Average	83	81	80	83	81	83

case the above conclusion is wrong. As will be shown shortly there is a heavy selection in this group, favoring the better quality of high school students not taking chemistry but taking both Latin and French.

Another conclusion also appears. Students taking Latin or French in high school are found to do better in college than those who do not. Here again the final conclusion must be modified by the quality of the students taking the two groups.

No particular difference can be noted between the training value of either high school Latin or French. Such being the case it is a fair question to ask if the utilitarian value of French as a modern language does not merit a wider consideration of French as the beginning language to the student in high school rather than Latin.

The data on the high school records of these groups will evaluate the degree of selection which has taken place in the mental quality of the students composing them. From these data proper allowance may be made for the effect of this selection. Table XIII gives these data.

The results for this table are shown graphically in Fig. 3.

Examination of Fig. 3 shows that it bears a striking resemblance to Fig. 2. The students who took Latin and French are consistently slightly ahead of the students that did not. The selection here is obvious. For chemistry, the students who took it in high school were slightly less proficient than those that did not. These same

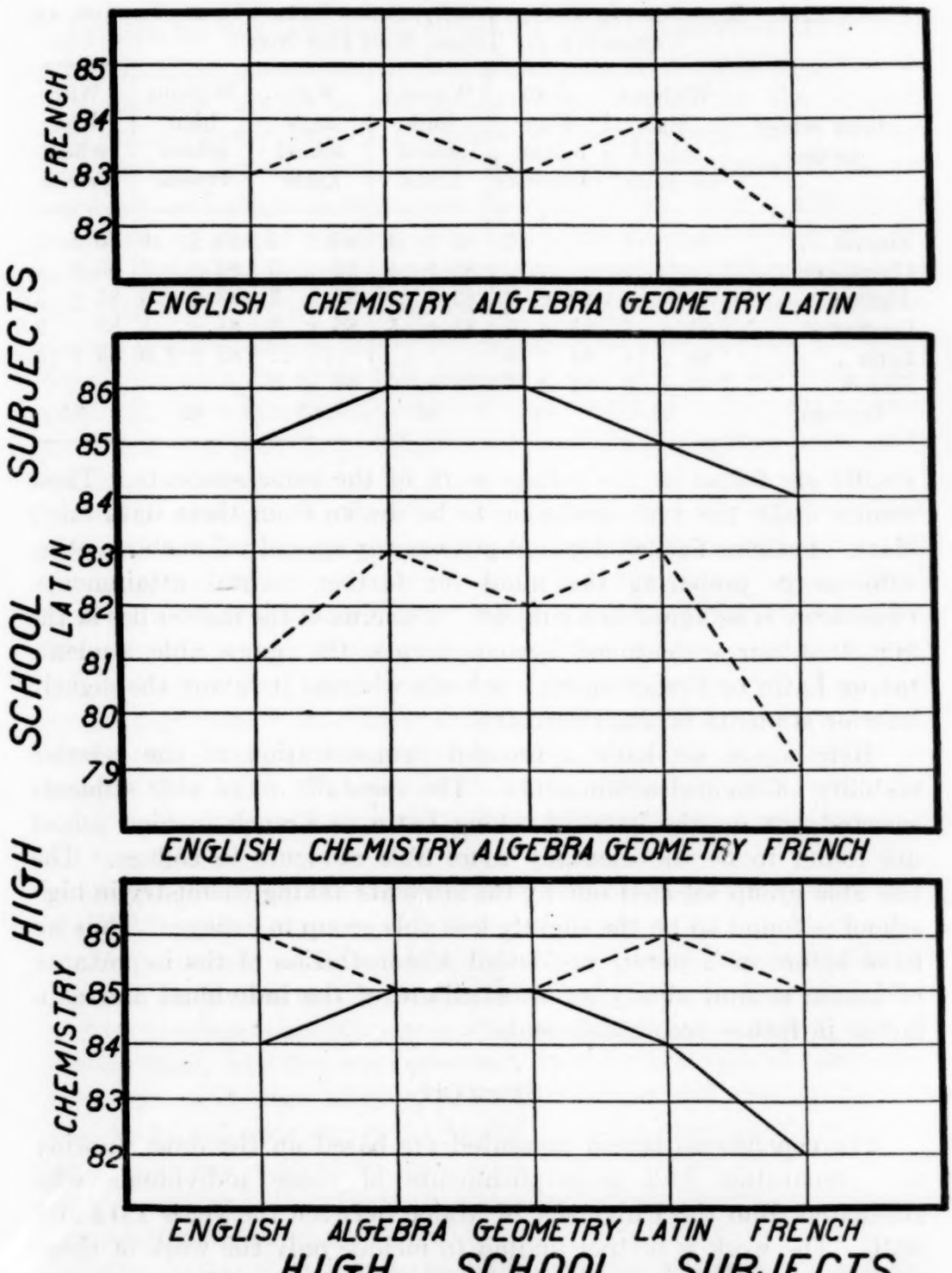


FIG. 3.—Graphs showing the high school ranks of those students who studied chemistry, Latin, and French in high school *vs.* those who did not. The solid line represents those who took the subject, the dotted line those who did not.

TABLE XIII.—HIGH SCHOOL GRADES FOR STUDENTS TAKING LATIN, FRENCH, AND CHEMISTRY *vs.* THOSE WHO DID NOT

High school subject	Without high school chemistry	With high school chemistry	Without high school Latin	With high school Latin	Without high school French	With high school French
English.....	86 ± .4	84 ± .2	81 ± .4	85 ± .2	83 ± .6	85 ± .2
Chemistry.....	.....	.....	83 ± .6	86 ± .3	84 ± .7	85 ± .3
Algebra.....	85 ± .4	85 ± .3	82 ± .5	86 ± .3	83 ± .6	85 ± .3
Geometry.....	85 ± .5	85 ± .3	83 ± .5	85 ± .3	84 ± .8	85 ± .3
Latin.....	86 ± .4	84 ± .3	.....	.....	82 ± 1.0	85 ± .3
French.....	85 ± .5	82 ± .3	79 ± .6	84 ± .3		
Average.....	85	84	82	85	83	85

results are found in the college work of the same students. These results make the real conclusion to be drawn from these data fairly clear. Latin or French does not possess any special value above other subjects in preparing the mind for further mental attainments. Chemistry is as valuable a subject. The crux of the matter lies in the fact that our educational system favors the more able students taking Latin or French in high schools whereas it favors the slightly inferior students taking chemistry.

Here again we have a forceful demonstration of the relative stability of mental attainments. The mentally more able students selected out on the basis of taking Latin or French in high school are found to be the mentally more able students in college. The less able group selected out by the students taking chemistry in high school is found to be the slightly less able group in college. Thus we have before us a purely accidental demonstration of the importance of innate mental ability as an attribute of the individual and as a factor in future accomplishments.

SUMMARY

The conclusions herein presented are based on the data showing the preparation and accomplishments of those individuals who graduated from the University of Maine between the years 1913 and 1921. The work is further limited to include only the work of these students during their Freshman year (that is 1909–1917) in conjunction with their ranks attained in preparatory schools. The data are extensive, involving 927 individuals. High school attainments

have been limited and measured by the work of the students in high school English, chemistry, algebra, plane geometry, Latin, and French. The students' ability in college has been studied for the subjects English, chemistry, algebra, analytical geometry, French, and German. The authors are aware that the data have weaknesses—these are admitted and some discount has been made for them. The weaknesses are of a kind which do not invalidate the conclusions, but are rather such as make the conclusions conservative. The analysis of the data leads to certain conclusions bearing more particularly on the human intellect and its expression in high school and college. Despite the slight attention given this subject in the past, it is the foundation on which any educational system must be built as well as the support to which any previously established educational system must rally if it is to be maintained.

Study of high school ability in comparison with college attainments shows that marks obtained by college students are in a measure dependent on the ranks obtained by the students in high school. This interrelation, however, is not so great as might be supposed. Thus, the average weighted correlation coefficient between high school English and the seven college subjects analyzed is .261. Between high school chemistry and the college subjects this weighted coefficient is .247; with high school algebra it is .294; high school geometry, .278; Latin, .271; with elementary French in high school it is .336; advanced French, .338. The differences of the correlations for the different subjects are not significant. That is, the better than average high school student becomes, in general, better than average in all college subjects.

From the above figures it will be noted that high school English is apparently just as good an indicator of the student's subsequent ability in college chemistry as is chemistry itself. Identity of subject-matter then, does not seem necessary to make previous training effective. As to identity of procedure, chemistry would seem to be more like analytical geometry than English, yet the correlation coefficients show that one is about as good an indication of the ability in analytical geometry as the other. So this does not seem sufficient to explain the relations as they do exist.

The study of the presence or absence of given courses of high school study brings out the fact that all of the generally accepted high school subjects have approximately equal training value. Any difference in training value appears to be due to the unconscious

selection of proficient students who take or do not take certain subjects. Thus with Latin or French, students who are, on the average, two points better in their other high school work take these subjects. The better students seem to avoid high school chemistry. These same innate biological differences again find their expression in the subsequent work of these students in college. Mental attainments are attributes of the individual capable of demonstration by the maintenance in after life of the position established in the younger years. As subsequent papers will show the relative degree to which this position is maintained appears to be due to innate intelligence and not to environment.

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

REPORTED BY C. O. MATHEWS
Graduate Student, Teachers College
Columbia University

Intelligence Testing

Army Alpha Scores in Relation to Progress in High School and to Mental Growth. Robert S. Ellis. School and Society, Oct. 3, 1925, 439–440. The intelligence of high school students who graduate and those who do not graduate in their class is compared. Age and grade medians are given.

Psychological Tests in Guidance. Morris S. Viteles. School and Society, Sept. 19, 1925, 350–356. This is a report by a committee of the Vocational Guidance Association of Philadelphia and vicinity setting forth ten principles which they consider fundamental to the scientific use of tests in guidance work.

An Experiment in Homogeneous Grouping. Ellen A. Maher. The Journal of Educational Method, Sept., 1925, 25–28. A plan for classification, making use of intelligence and educational tests and estimates of moral, social and physical traits.

The Relation of Ability to Achievement. W. H. Pyle. School and Society, Sept. 26, 1925, 406–408. Data give the relationships between intelligence tests and measures of achievement. Factors influencing achievement discussed.

Achievement Testing

The So-Called Guessing Element in Written Tests. Public Personnel Studies, Oct., 1925, 275–282. The newer types of tests are shown to be influenced much less by the guessing element than the free answer form.

Making a Local Composition Scale. Earl Hudelson. The School Review, Oct., 1925, 601–609. The explanation of a procedure for constructing valid local composition scales.

New York State Regents Literacy Test. J. Cayce Morrison. Journal of Educational Research, Sept., 1925, 145–155. The history and underlying principles of the New York State Regents Literacy Test. A sample test is given.

A Study of the Relation of Reading Ability to Scholastic Achievement. Charles E. Dickinson. The School Review, Oct., 1925, 616–626. Three standard tests were used to measure the reading ability of high school students and these measures were correlated with teachers' marks.

The Experimental Study of Education with Special Reference to the Elementary School. Paul Klapper. Journal of Educational Research, Sept., 1925, 123–135. A discussion of factors which are retarding the development of objective measures.

Learning and Psychology of School Subjects

A Statistical Study of Children's Errors in Sentence-Structure. S. L. Pressey. The English Journal, Sept., 1925, 529–535. A study of errors in 980 composition

papers of pupils in Grades VII–XII. Frequencies of errors are given and findings of other investigators are compared.

The Relation of Class Size to Control of Attention. Lofter Bjarnason. The Elementary School Journal, Sept., 1925, 36–41. The author concludes that group attention is affected more by the technique of teaching employed than by the size of the class.

A Silent Reading Experiment. Ellsworthy Lowry. The Elementary School Journal, Sept., 1925, 42–49. Ten weeks intensive training in silent reading showed an average gain in speed greater than that usually made in three years, and an average gain of 25 per cent in comprehension.

The Effect of Emotional Disturbance on Ability to Concentrate. R. Clyde White. School and Society, Sept. 12, 1925, 343–344. The effect of the excitement accompanying "yell practice" upon ability to solve arithmetical problems is tested.

Insight and Skill in Arithmetic. E. R. Hamilton. Journal of Educational Research, Sept., 1925, 136–144. A criticism of some past researches in arithmetic with three suggestions for improvement.

Character and Personality

The Incidence of Undesirable Behavior in Public School Children. M. E. Haggerty. Journal of Educational Research, Sept., 1925, 102–122. An attempt at a technique for identifying and giving objective description to socially undesirable behavior in school children.

Personality Differences in Intelligence Test Performance. Grace E. Mason. The Journal of Applied Psychology, Sept., 1925, 230–255. An attempt is made to measure cautious reactions by a detailed study of errors and omissions on an intelligence test. Omissions are shown to be controlled by numerous factors.

Personality and Vocational Achievement, Part II. George C. Brandenburg. The Journal of Applied Psychology, Sept., 1925, 281–292. This is the report of a follow up study of 36 college men five and a half years after graduation, showing the influence of personality and character factors on success.

The Reliability and Relationships of the Colgate Mental Hygiene Test. Ralph K. Hoitsma. The Journal of Applied Psychology, Sept., 1925, 293–303. This test is an extension of Woodworth's "Psychoneurotic Inventory."

A Study of the Literature on the Downey-Will-Temperament Test. Leo A. Kuntz. The Catholic Educational Review, Oct., 1925, 478–485. A paper on the validity and reliability of the Downey test, in the light of experimental work reported upon it.

Personality Tests: White and Negro Adolescents. Dagny Sunne. The Journal of Applied Psychology, Sept., 1925, 256–280. White and negro children are compared on the basis of responses to three tests: the Pressey X-O tests, the Carnegie IX and the Woodworth-Mathews questionnaire.

Statistical Methods

Job-analysis Sheet for Computing Partial and Multiple Coefficients of Correlation and Regression Coefficients. Percival M. Symonds. Teachers College Record, Sept., 1925, 52–67. The author describes the process of computing coefficients by the use of a correlation chart.

Exceptional Children

Phenomenal Memory in Its Bearing upon Various Mental Tests. Francis E. Otis. The Journal of Applied Psychology, Sept., 1925, 311–318. A case report of a feeble-minded boy with a remarkable memory.

Where Gifted Children Learn. Paul Hanley Furfey. The Catholic Educational Review, Oct., 1925, 465–472. A case study of a gifted boy. Undirected reading, play and observation are discussed as sources of knowledge for gifted children.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

CONDUCTED BY JOHN HOCKETT
The Lincoln School of Teachers College

A Text in Educational Psychology

Elements of Educational Psychology, by Lawrence A. Averill. Houghton Mifflin Co., Boston, 1924. Pp. XII + 425.

This is a much more pretentious volume than Dr. Averill's first venture in the psychology text field, his *Psychology for Normal Schools*. "It is intended primarily as a textbook for teachers-in-training who have had a general introductory course in psychology." There are three parts to the book: I, an introductory discussion of the Motivation of Learning; II, an extended and detailed treatment of the Learning Process; and III, a brief consideration of the Results of Learning. An appendix outlines the nature and method of modern measurement.

The method of instruction defined by the book comprises experiments of simple, well-known, laboratory kinds, surveys of the classic studies, statements and applications of laws and principles, and the solution of practical problems. The author accepts the classification of learning into sensori-motor, perceptual, associative, and problem types. He divides the learning process into initial and secondary phases: the first concerned with impressing the data and, therefore, conditioned by attention, learning capacity, and fatigue; the second concerned with retaining the data and, therefore, conditioned by memory, forgetting, ideation, and training. The transference of training is to him, as to many, a matter about which we still know little and about that little agree less. His advice to teachers is not to miss the possibility of the general by slighting the specific.

The author's style might have been less discursive, his diction at times might have been less florid, and his philosophy might have been kept more clearly separated from his psychology; but the measure in which the contraries exist is not damaging. On the whole the book is a full one, a scientifically acceptable one, and a well-written one.

M. H. Willing.

A New Text in Educational Psychology

Education as the Psychologist Sees It, by W. B. Pillsbury. The Macmillan Company, New York, 1925. Pp. VII + 342.

It would be useless, probably, to debate the question of when educational psychology becomes psychologized education. However, the author seems to assume that the two phrases are identities; and by calling his book a text in education he has given it a label of a difference which it does not seem to possess. There is a professional seriousness, however, in the turn of the name. In it we find frank recognition of the prevalent fact that much of our educational psychology, so-called because it is supposed to be applied and hence vocational, still remains almost as academic and as little vocational as are standard introductory courses in general psychology. This ought not be true. Unless their understanding and teaching techniques have been much improved by such, requirements that teachers must have had courses in educational psychology (or any courses) cannot be wholly justified. It is a usable psychology which thoughtful teachers are demanding. And for utility teachers of elementary and secondary schools require not abstract generalizations but concrete applications. For such a presentation whenever it may appear, for it will sometime be written, it is suggested that the name psychologized education be preserved.

But so far as the book is concerned the question about its name is trivial. The volume is an interesting and effective presentation of the facts of elementary educational psychology. Chapter topics such as distribution of traits, original nature, individual differences, aims of education, habit, attention, learning, reasoning, feeling and emotion, will, fatigue, transfer, and objective tests indicate its conventional subject-matter. However it is not ordinary in style and presentation; it is superior. Treatment of topics is direct, simple, and wholesome. In an enviable style the author has dodged dogmatic for wholesome statements and has not become involved in qualifications. A splendid text has been prepared. Among recent books in educational psychology this one is predestined to gain and to hold a very favorable place as a text in elementary courses. For his achievement the author is cordially congratulated.

Edwin Maurice Bailor.

Dartmouth College.

Principles Underlying Moral Guidance

Formative Factors in Character, by Herbert Martin. New York: Longmans, Green and Co., 1925. Pp. 346. Cloth $1.40.

Teachers, students of education, parent-teacher associations, reading circles, and study clubs interested in the moral development of children will find in this book a full discussion of the principles underlying moral guidance. Attention is constantly directed to the adult need of understanding the child mind and its meanings, not as a "little man," but as a growing process. The moral self is a product of age, experience and its interpretation and continuous reorganization. The author reviews the psycho-biological equipment for receiving and evoking moral experience, discusses the influences of bodily conditions on conduct, considers the conditioning power of biological and social inheritance in the development of personality, analyzes the relation of such psychological factors as thinking, desire, interest, will, suggestion, and imitation to moral conduct, and stresses the social value of play and redemptive power of work. Several chapters are devoted to the possibilities of conduct control by such institutions as the home, the school and various non-professional agencies. Emphasis is particularly laid on the part of adult responsibility in defining clearly worthy ends and purposes to which the children may be won, and in giving opportunity for these ideals to function in concrete situations. The educational problems incumbent on a democracy are considered from various angles. The closing chapter envisions religion as "morality, conscious of its social significance, seen in its cosmic setting." Morality is summed up as "reverence for sacred things, respect for personality, the will to friendship, the spiritual nature of man, the reality of the Unseen, and an experimental verification of the supremacy of a life of righteousness motived by love."

Gladys C. Schwesinger.

The Elements of Statistics and Measurement

Beginnings in Educational Measurement, by Edward A. Lincoln. Philadelphia: J. B. Lippincott and Co., 1924. Pp. 147.

This is an easy book intended for persons who have little background, meager facilities for reference and little opportunity for personal contact with the instructor. It was written, as the author informs us, "for the purpose of assemblying for the teacher in service a body of information which will enable him to think and talk intelligently about educational measurement."

In a simple, clear, "talky" manner the author briefly discusses these questions among others: characteristics of a standard test; measures of central tendency; variability; reliability; measurement of intelligence; the measurement of character. Of course, none of these problems is given a thorough treatment. This book is not intended for and will not serve graduate students who want to specialize in statistical procedure. It does no more than touch upon the more important phases of educational measurement. But the popularizing is well done. It should fulfill its limited but important purpose.

H. MELTZER.

Oregon State College.

THE TRAINING OF HIGH SCHOOL TEACHERS

Professional Secondary Education in Teachers Colleges, by Alfred Lawrence Hall-Quest. Teachers College, Columbia University, New York, 1925. Pp. VIII + 125.

On the basis of courses offered and time allotments, which institution best prepares high school teachers for their work, the university or the teachers college? Analysis of catalogs, correspondence with institutional authorities, and personal visitation lead the author of this doctor's dissertation to the following conclusions. The courses offered in both types of institutions for the prospective high school teacher are listed under a bewildering diversity of names. The universities emphasize educational theory and general principles of educational science; the teachers colleges, practical methods. The universities emphasize cultural background; the teachers colleges, special courses and curricula. "Teachers colleges prepare future teachers to teach; universities educate them into more comprehensive scholarship."

Dr. Hall-Quest offers, as his solution of the problem raised by this divergence, the development in both institutions of "professional courses whose content is an integration of all factors essential to an expert teacher in a particular field." The trend toward such courses is already marked in the teachers colleges.

A comparison of teachers colleges and universities which takes account only of courses and curricula in catalogs is, of course, very incomplete. One would like to know also how these institutions compare in quality of instruction, educational equipment, scholastic standards, character of student body, and subsequent success of graduates. This investigation, however, limited as it is, does assemble

valuable data for the support of what seems to have been the author's *a priori* thesis; namely, that there is great need of a standardization of teacher-training and that this may best be effected by a cooperative development of well coordinated, professionalized subject-matter courses.

M. H. WILLING.

SOURCE BOOK IN SECONDARY EDUCATION

Principles of Secondary Education, by Willis L. Uhl. Silver Burdett and Co., New York, 1925. Pp. XII + 692.

This large volume is calculated to reduce by many hours the dreary, though highly competitive, library search of the student in secondary education for illusive single copies of perennially nourishing periodicals, surveys, special reports, bulletins, pamphlets, circulars, yearbooks, and monographs. Over a hundred significant excerpts from these exceedingly valuable but unduly inaccessible forms of professional literature are gathered together and presented here in an orderly way. They cover such topics as The American Plan of Secondary Education, The Secondary School Teacher, Secondary School Pupils, Secondary Education in Foreign Countries, The Reorganization Movement in Secondary Education, and Curriculum Problems. In each chapter Professor Uhl has provided brief and pointed introductions, transitions, and summaries,—the latter in the shape of lists of principles derived from the quoted material. He also has furnished frequent lists of study and discussion suggestions, and extensive bibliographical notes.

Such a compilation at the present time must, of course, be based on the personal judgment and likings of the compiler. Doubtless it would not be difficult for another professor of secondary education to point out serious omissions from it or serious over-emphases. Any teacher, however, could not fail to find a great deal here of which he approved and with which he would wish his pupils to be familiar. It is the sort of thing much needed among pedagogical helps to-day and might profitably be duplicated in a half dozen other fields besides the secondary. Students in training schools often feel that the hunt for assigned references exhausts all but hypothetical disciplinary values long before it terminates successfully. They will applaud the appearance of fat source books like this, whether the producers of them act from sympathetic motives or from eagerness to get a larger amount of reading satisfactorily and promptly completed.

M. H. WILLING.